DECOUPLED MULTIVARIATE TIME SERIES MODELS FOR MULTISITE STREAMFLOWS

by
M. KRISHNASAMI

DEPARTMENT OF CIVIL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY KANPUR

JUNE, 1976

DECOUPLED MULTIVARIATE TIME SERIES MODELS FOR MULTISITE STREAMFLOWS

A Thesis Submitted
in partial Fulfilment of the Requirements
for the Degree of
DOCTOR OF PHILOSOPHY

by
M. KRISHNASAMI

to the

DEPARTMENT OF CIVIL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY KANPUR
JUNE, 1976

TO MY BELOVED FATHER

WHO HAVING AWAITED PATIENTLY ALL NIGHT

COULD NOT LIVE TO SEE THE STREAK OF DAWN

CERTIFICATE

This is to certify that the thesis 'Decoupled Multivariate Time Series Models for Multisite Streamflows' submitted by Shri M. Krishnasami, in partial fulfilment of the requirements for the Degree of Doctor of Philosophy at the Indian Institute of Technology, Kanpur is a record of bonafide research work carried out under my supervision and guidance. The work embodied in this thesis has not been submitted elsewhere for a degree.

(S. RAMASESHAN)
Associate Professor
Department of Civil Engineering
Indian Institute of Technology
Kanpur

Dated: June 24, 1976

ACKNOWLEDGEMENTS

The author is deeply indebted to his thesis adviser, Professor S. Ramaseshan who suggested the problem of this thesis and gave invaluable guidance and encouragement at every step of investigation.

Acknowledgement is due to the Computer Centre, IIT Kanpur and to the Computer Centre, IIT Madras for allowing the use of the computing facilities. The author is thankful to the authorities of Annamalai University for granting him the necessary study leave.

Thanks are due to Messers Kuganathan, Shyamal Ray, Rajagopalan, Krishnan and Vadivelu for their kind help, to Shri J.K. Misra for the efficient typing and to Shri Verma for the preparation of drawings.

The author owes deeply to his wife Uma who shared with him the agonies and ecstasies besides rendering signal secretarial assistance; and to his parents-in-law and mother, and lastly to his father (albeit posthumously) whose concern and interest in this venture transcend description.

M. Krishnasami

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure No.		Page
4.16	Seasonwise Correlation of Tendaily Residuals of Stationary AR Models; River 3	77
4.17	Variance of Monthly Residuals of Stationary First Order AR Model	80
4.18	Variance of Tendaily Residuals of Stationary Third Order AR Model	81-82
4.19	Seasonwise Correlation of Monthly Residuals of Nonstationary First Order AR Model	92-93
4.20	Seasonwise Correlation of Tendaily Residuals of Nonstationary First Order AR Model	94-95
4.21	Seasonwise Correlation of Tendaily Residuals of Nonstationary Second Order AR Model; River 1	95
4.22	Seasonwise Correlation of Tendaily Residuals of Nonstationary Second Order AR Model; River 2	96
4.23	Seasonwise Correlation of Tendaily Residuals of Nonstationary Second Order AR Model; River 3	96
4.24	Seasonwise Correlation of Tendaily Residuals of Nonstationary Third Order AR Model; River 1	97
4.25	Seasonwise Correlation of Tendaily Residuals of Nonstationary Third Order AR Model; River 2	97
4.26	Seasonwise Correlation of Tendaily Residuals of Nonstationary Third Order AR Model; River 3	98
4.27	Variance of Monthly Residuals of Nonstationary First Order AR Model; River 1	98
4.28	Variance of Monthly Residuals of Nonstationary First Order AR Model; River 2	99
4.29	Variance of Monthly Residuals of Nonstationary First Order AR Model; River 3	99

LIST OF SYMBOLS

[A]	Lagone coefficient matrix in Matalas model
AR	Autoregressive
ARIMA	Autoregressive - integrated - moving average
ARMA	Autoregressive moving average
B	Vector of parameters of ARMA model
b	Autoregressive coefficient in Thomas-Fiering model
[C]	Lagone coefficient matrix in decoupled multivariate model
CL	Confidence limits
cpy	Cycles per year
C(t)	Partial sum of flows upto time t
[D]	Lagzero coefficient matrix in decoupled multivariate model
DOF	Degrees of Freedom
d	Degree of differencing
E	Expectation operator
E(t)	Standardised univariate residual vector at time t
est	Estimate
e(t)	Standardised univariate residual at time t
F	F statistic, cumulative distribution function
FGN	Fractional Gaussian Noise
f(i)	Observed frequency for the i-th class
Fig	Figure

$G(u)$	Spectral ordinate at the u-th frequenty
g	Coefficient of skewness
H	Hurst coefficient
I	Number of classes in frequency analysis
i	Subscript
J	Number of parameters in a distribution
j	Superscript denotes season j
K	Total number of sites; Chow's frequency factor
l	Serial number of the site; subscript
MA	Moving average
M_o	Lagzero correlation matrix
M_{-1}	Lagone correlation matrix
MLM	Method of Maximum Likelihood
MLS	Method of least squares
MM	Method of moments
m	Variable in F-statistic
N	Total number of data points
p	Order of autoregression
PD	Positive definite
$p(i)$	Probability that a variable belongs to the i-th class
$Q(t)$	Streamflow at time t
q	Order of moving average
R	Range
r	Order of the sample moment

r_i	Estimated correlation coefficient at lag i
S(B)	Sum of squared errors in ARMA model
s	Number of seasons
s_x	Standard deviation of x
SE	Standard error
T	Total time; theoretical
t	Time, t - statistic
Var	Variance
V(B)	Variance - Covariance matrix of parameter estimates in ARMA model
x(t)	Value of normalised standardised series at time t
$\bar{x}$	Mean of the x series
y(t)	Normalised parametrically standardised value at time t
$\bar{Y}$	Mean of the y-series
Z	Z statistic
Z(t)	Normalised standardised value (parametric) at time t
α	Level of significance
∇	Operator for differencing
$\varepsilon(t)$	Univariate residual at time t
$\eta(t)$	Multivariate residual at time t
μ	Population mean
μ_n	Mean of the logarithms of the series
ρ_i	Correlation coefficient at lag i
$\sigma(n)$	Standard deviation of the logarithm of the series

σ_y	Population standard deviation of y
θ_i	Moving average coefficient of i-th order
ϕ_i	Autoregressive coefficient of i-th order
ϕ_{ii}	Partial autocorrelation coefficient at i-th lag
χ^2	Chisquare statistic

SYNOPSIS

DECOUPLED MULTIVARIATE TIME SERIES MODELS FOR MULTISITE STREAMFLOWS

A Thesis Submitted
In Partial Fulfilment of the Requirements
For the Degree of
DOCTOR OF PHILOSOPHY
by

M. KRISHNASAMI
to the
Department of Civil Engineering,
Indian Institute of Technology, Kanpur, India

The design and operation of a water resources system should be based on a clear understanding of the processes affecting them which in turn depend on proper analysis and interpretation of data. Particularly important are the streamflow data which exhibit complex stochastic characteristics. They include: 1. a nonstationary behaviour in terms of the seasonal variation of the mean, standard deviation, serial dependence, etc.; 2. trends due to natural or human influences; 3. periodic or cyclic components; 4. autoregression or persistence within the time series; 5. moving average or external correlation between random components at the site; and 6. correlation in space or among time series. Generally factors 1 to 5 are studied in the univariate modelling of a time series, say, the streamflow at a site and factor 6 is studied in the multivariate modelling of several time series. Knowing the dependence or independence of streamflows at different sites,

it is possible to design and operate multireservoir systems by taking advantage of the interdependence.

Mathematical modelling of streamflows at a site is itself fairly complicated. Modelling of multisite streamflows is further complicated by the fact that the multivariate streamflow variability may be attributed to i. the within-the-station variations in terms of factors 1 to 5, and ii. the among-the-stations variations in terms of the crosscorrelations between the streamflows at different sites. Univariate models in hydrology deal with the former and multivariate models generally consider both simultaneously. It seems possible to decouple the variability due to these two classes so that the within-the-station variability can be first estimated by univariate modelling and used for identifying the serially independent random components at each site; and then these random components can be related through a multivariate model to account for the inter-station correlations and the among the stations variability. Such an approach seems to be advantageous because of the following:

1. The expertise developed in the past in univariate time series modelling can be beneficially used.

2. As the serially independent and normally distributed variates are used in the multivariate model, a better performance of the model can be expected.

3. Since the parameters are estimated in stages, the number of parameters estimated in each stage is comparable to that of univariate and multivariate model respectively and the problem of simultaneously estimating all the parameters is avoided.

4. Where data lengths vary from station to station, all the available data can be used in the estimation of paramdters and decoupling may eliminate errors and bias that may be present in the simultaneous estimation of parameters.

5. Decoupling of spatial and temporal variations has been suggested recently by Yevjevich, Mejia and Iturbe, and Yevjevich and Karplus. Rather than using simple spectral and correlation procedures to represent spatial variability, the use of multivariate models is suggested in this study. This is similar to the approaches of Frost and Clarke, and Yevjevich.

6. Decoupled multivariate models seem to be very powerful tools in representing complex multiple interrelated time series.

Without loss of generality, the study was restricted to three sites in a river basin in North India for which data were available and to a 25 year concurrent record of streamflow. Data series considered in this study include i) the

tendaily historical data series; ii) the monthly data series calculated from the tendaily series as the average of the three tendaily values in each month; and iii) the annual series, being the average of the monthly values in each year. Annual data were normally distributed; but monthly and tendaily series were normalised by logarithmic transformation. The data series were standardised to a mean zero, unit variance series by means of the sample means and standard deviations for each of the periods. Using ARIMA models and a nonlinear least squares regression approach, univariate models were fitted to the standardised normalised time series at each site. The results indicated autoregressive models of orders zero, one and three respectively for the annual, monthly and tendaily series. The residuals were calculated and tested for serial independence. Time domain and frequency domain analyses indicated serial independence when the full series was considered; but, for monthly and tendaily series the correlation coefficients between the residuals of adjacent periods were found to be significant in several cases. This indicated a seasonal variation of persistence and so nonstationary autoregressive models of the same order as before but with seasonally varying parameters were fitted to the monthly and tendaily series. The residuals from the nonstationary models were found to be serially uncorrelated seasonally and as a whole. The univariate

residuals were also generally normally distributed.

The second stage consists in fitting a multivariate time series model to the univariate residuals to take into account the among-the-series variation. Using currently available procedures (Matalas, Young and Pisano), a stationary first order autoregressive multivariate model was fitted to the univariate residual series. The multivariate residuals were evaluated and tested for independence serially and with one another. When considered as a whole, the series were found to be independent serially and of one another; but monthly and tendaily series showed significant cross-correlation on a seasonal basis in several periods. This suggested a nonstationary multivariate model with seasonally varying parameters. For the tendaily series, for four seasons it was not possible to fit nonstationary first order autoregressive multivariate models using the procedures adoptedffor stationary modelling. This is because of inconsistencies in the sample correlation matrices and the problem was similar to cases mentioned by Matalas and Wallis. Zerooth order models were fitted to the tendaily series in such cases. Development of procedures for fitting first order models and for estimation of parameters in such cases is beyond the scope of this study. Monthly data did not exhibit any such problems. The multivariate residuals from the nonstationary model were estimated and when tested statistically, they were

found to be independent serially and of one another, seasonally and as a whole.

It is possible to recouple the multivariate and univariate models and also use inverse transformations to denormalise and destandardise them in the proper order so that the complex relationships between the actual flows can be derived in terms of the parameters estimated earlier. These relationships can be used to predict missing data if any; forecast future values of the streamflows at any site; and generate multisite streamflow data for simulation of complex water resources systems.

One step ahead forecasting and predictions were considered in this study using the multivariate model developed earlier. It was seen that generally multivariate estimation and forecasting leads to a better definition of the expected value and confidence levels of the variable concerned than in the case of univariate models.

On the basis of this study, the following conclusions can be made:

i) Decoupling facilitates the use of simpler component models in the modelling of complex stochastic processes, and leads to a better understanding of the temporal and spatial dependence between streamflows at different sites.

ii) For the standardised annual series, a zeroeth order univariate autoregressive model and a first order multivariate autoregressive model are indicated. For the normalised, standardised monthly series, a nonstationary first order autoregressive univariate and multivariate model are indicated. Tendaily data after normalisation and standardisation, can be represented generally by a third order autoregressive nonstationary univariate model and a first order autoregressive nonstationary multivariate model. For some seasons, the sample correlation coefficients for the nonstationary model were inconsistent with the assumption of a multivariate first order autoregressive model. In such cases, a zeroeth order multivariate model was fitted to the tendaily series.

iii) Multivariate models can be used for the estimation of missing values, and for one step-ahead forecasting. They generally lead to a better definition of the conditional expectation and a smaller standard deviation of the forecast or estimate.

iv) Data can be generated on the basis of the decoupled multivariate model and use in simulation analysis of multi-reservoir systems. Decoupling is a very powerful tool in the mathematical modelling of interrelated time series and the procedure developed herein can be used for modelling other multivariate processes.

1. INTRODUCTION

1.1 General

The design and operation of a water resources system should be based on a clear understanding of the processes affecting them which in turn depends on a proper analysis and interpretation of data. This problem is becoming increasingly important owing to the necessity of optimally utilising the limited available resources of the earth particularly because population and per capita demand are increasing exponentially. Historically a number of procedures have been developed and applied by hydrologists and water resources engineers to quantify and evaluate the availability and variability of water resources in time and space. These include (i) mass curve and related analyses of historical data (ii) empirical correlation coefficients (iii) deterministic simulation models and (iv) stochastic models. They will now be dealt with briefly.

1.1.1 Mass curve and related analyses of historical data

The traditional method of estimating the required storage capacity of reservoirs makes use of the historical data of streamflows. Using the mass diagram of inflows and

demands, minimum storage required is estimated under the condition that no water shortage occurs for the historical sequence of inflows. A modification of this method is the sequent peak algorithm (Fiering, 1967). This is a computer-oriented numerical procedure which makes the estimates of storage less ambiguous than those obtained by the graphical procedure. But both procedures generally make use of only the historical record of streamflows and it is unlikely that the same sequence will be repeated during the life of the structure.

1.1.2 Empirical correlation relationships

These relationships are used to estimate runoff when historical data of runoff are inadequate. They make use of the correlation of runoff with independent variables like hydrometeorological and catchment characteristics. A set of three or more variable charts that relate the dependent variable in terms of the independent variables is prepared. For example, Linsley's coaxial chart (Linsley et al., 1958) relates runoff to precipitation, antecedent moisture and week of the year. Such graphical procedures are useful when independent variables are nonlinearly correlated among themselves and the nature of relationship is not known.

1.1.3 Deterministic simulation models

Precipitation forms the input to the hydrologic basin and, runoff, both surface and subsurface, forms the output. Runoff is obtained by the modification of precipitation by evaporation, infiltration and surface and subsurface storages. An estimate of runoff can be made on the basis of quantitative assessment of these component processes. From known concurrent data of precipitation and other hydrometeorological data as well as runoff data and on the basis of an assumed mathematical model for the basin, the parameters of the model are estimated. Knowing the parameters and for any given or assumed precipitation the component processes including runoff are estimated through simulation. The simulation models in use include (i) Stanford Watershed model (ii) The Utah Watershed Simulation model (iii) The Agricultural Research Service model (iv) The Columbia River Basin model. Details of these models are available in references (Nordenson, 1969: Clarke, 1973).

1.1.4 Stochastic models

Stochastic process models: Stochastic process models in hydrology include simple bivariate models (Sariahmed et al., 1968), models of finite duration processes (Ramaseshan, 1971), queuing models (Langbein, 1958), transition probability models (Moran, 1959), random walk models (Cox et al., 1965) and

linear stochastic system models (Rao et al., 1971). They use mathematical theory to study the processes and to make inferences concerning the characteristics of the processes. Because only simpler mathematical models can be analytically solved, they are useful only for simple idealised representations. Real hydrologic systems are nonlinear and interacting. Furthermore hydrologic processes are highly seasonal and they also exhibit serial- and crosscorrelation that may vary seasonally. Hence analytic stochastic models listed above are of limited use in studying the stochastic characteristics of streamflow.

Time series models: A sequence of values occurring or observed along time is called a time series. Unlike frequency analysis where the serial dependence between events is ignored, time series analysis takes the sequence of occurrence of events into account. Time series may in general have the following components: (i) Trend: This indicates the long term deterministic change in the time series and may be due to natural or human agencies, (ii) cyclicity: This refers to the periodic component that repeats itself at definite intervals. Annual cycle is an example of cycles that are important in streamflow series; (iii) autoregression (AR): This refers to the linkage existing between the value at a given time with earlier values of the same series,

(iv) moving average (MA): This is the external correlation between the flow and the present and past random components at the site, and (v) the pure random component. Generally these five factors are considered in the univariate modelling of a time series, as, for instance, the time series of streamflows at a site. There is, in addition, the external persistence which refers to the linkage between the variable at a given time and concurrent as well as earlier values of other time series. These are considered in the multivariate modelling of several time series. Knowing the dependence or independence of streamflows in a multireservoir system, it is possible to design and operate the system better.

1.2 Statement of the Problem

Multivariate modelling involves the representation in mathematical terms of the variabilities of streamflow at each site and among the sites. This is however complicated by the fact that the variability of multisite streamflows may be attributed to (i) the within-the-station variation, say, in terms of the AR, MA and the serially independent random components at each site, (ii) the among-the-station variation in terms of the crosscorrelation between the streamflows at different sites, and (iii) the internally and externally independent pure random component at each site. In several earlier studies (Beard, 1965; Matalas, 1967;

Young et al., 1968; McMahon et al., 1972; Pentland et al.,1971; Matalas et al., 1971) they have been considered together. In this study it is proposed to develop a multivariate model that decouples the above variations.

The problem may be defined as follows:

i) to represent the multisite streamflows by decoupled multivariate time series models in which initially the within-the-station component is represented by appropriate univariate time series models; and subsequently the residual independent random components at each site that constitute 'prewhitened' time series are related through a general linear multivariate model in terms of the internally and externally independent pure random components;

ii) to test the validity of the model using available historical data, and

iii) to use the results for one-step-ahead forecasting.

1.3 Objectives of the Study

The objectives of the study are as follows:

i) to fit suitable univariate time series models to the streamflow at each site and hence to estimate at each site the serially independent random components;

ii) to fit a suitable multivariate model to the serially independent residuals at each site and hence to evaluate

the serially and mutually independent random components;

iii) to validate the fitted decoupled multivariate model;

iv) to identify the complex stochastic relationships between the multisite streamflows by inverting the relationships, and

v) to use the decoupled model for one-step-ahead forecasting.

1.4 Significance of the Study

The proposed approach for the analysis and stochastic modelling of streamflows at several sites seems to be advantageous because of the following reasons.

i) The expertise developed in the past in univariate time series modelling can be used beneficially.

ii) Since serially independent random normal variates are used in the multi-variate model, a better performance of the model can be expected than with the serially correlated data.

iii) Since the parameters are estimated in stages, the number of parameters in the two stages are comparable respectively to that of univariate and multivariate models and the problem of simultaneously estimating all parameters is avoided.

iv) Where data lengths vary from station to station, all the available data can be used in the estimation of the parameters

and decoupling eliminates errors and bias that may be present in the simultaneous estimation of parameters,

v) Decoupling of spatial and temporal variations has been suggested by Mejia and Iturbe (1974) and Yevjevich and Karplus (1972). Rather than using distance-correlation procedures, the use of a multivariate model is suggested to represent spatial variations. The concept is similar to that of Frost and Clarke (1973) and Yevjevich (1975).

vi) The procedure enables the most appropriate form and order of the univariate model to be chosen independently for each site. The form and order of the scheme for representing external variability are independent of the scheme for representing internal variability. This lends flexibility to the whole procedure enabling extraction of maximum information from internal and external correlation components. The decoupled multivariate model is hence a very powerful tool in representing complex multiple interrelated time series.

1.5 Scope of the Study

The scope of the present study is limited to the following:

i) Data of three tributaries to a major river in North India are taken up for analysis.

ii) The analysis is restricted to short lag models.

iii) The available data consist of (a) three values for each month, the first being the average of the flows of the first ten days of the month; the second, that of the next ten days and the third, that of the remaining days of the month. These data are referred to in this study as tendaily flows; (b) flows for the corresponding months are estimated by averaging the 3 tendaily flows for the month and (c) annual flows are computed by averaging the 12 monthly flows of the year.

iv) A first order AR multivariate model is used in this study to represent the relation between the univariate residuals.

v) Estimation of the parameters of the univariate and multivariate models is done according to some generally accepted procedures; and

vi) One-step-ahead forecasting is considered for the proposed model. Multistep ahead forecasting which is an extension of the above is not considered. Furthermore, forecasting and estimation are tested only for a few time intervals.

1.6 Organization of the Report

The study is reported in the following sequence:

i) Univariate and multivariate models for streamflow sequences are briefly reviewed. The proposed decoupled model for multisite streamflows is presented and compared with available multivariate models (Chapter 2).

ii) Details of the preliminary statistical analyses of data are reported in the following sequence: analysis of historical data with respect to its distribution and parameters; normalisation of the historical data using appropriate transformations and estimation of the parameters; and standardisation of the normalised data to have zero mean and unit variance (Chapter 3).

iii) Fitting of the univariate model to the normalised, standardised data is reported in terms of a general linear stationary stochastic model, viz., the autoregressive-integrated-moving average (ARIMA) model. Steps include identification of a suitable model; estimation of the parameters of the identified model; and diagnostic checking of the residuals for independence and normality. The analysis is extended to nonstationary univariate models (Chapter 4).

iv) A multivariate model is fitted to the serially uncorrelated residuals of the univariate model and the parameters

are estimated. The residuals are tested for normality and for internal and external independence. A nonstationary model is found to be necessary for monthly and tendaily data (Chapter 5).

v) The components are recoupled to yield the multivariate relationships. They are used for multivariate one-step-ahead forecasting (Chapter 6).

vi) A brief summary of the study is given, conclusions drawn on the basis of the results of the study are presented and some suggestions are made for future investigations (Chapter 7).

2. TIME SERIES MODELS IN HYDROLOGY

2.1 Introduction

Consider a stochastic process, say $Q(t)$, the streamflow at time t at a particular site. The sequence of values of the variable at times t_1, t_2, ... t_N, viz., $Q(t_1)$, $Q(t_2)$, ...$Q(t_N)$ represents a discrete time series referred to herein as the time series. The time series is a sequence of random variables with an associated probability distribution function for each value of t. Since a sequential relationship may be present between the values of the variable at different times it may be necessary to define the joint probability distribution functions associated with any arbitrary set of times, say, t_1, t_2, ...t_N. This concept provides a probabilistic description of the process. Time series models are used to define such relationships and time series analysis is used to model the process on the basis of a historical record which is only one of the infinite possible realisations of the underlying stochastic process.

A time series is said to be strictly stationary if the joint probability density function of the families of random variables $[Q(t_1), Q(t_2), \ldots Q(t_N)]$ and $[Q(t_{1+i}), Q(t_{2+i}), \ldots$ $\ldots Q(t_{N+i})]$ is the same for all i for any set of t_1, t_2...t_N;

i.e., stationarity is achieved if the distribution is invariant with respect to translation in time. This implies stationarity of moments of all orders of the time series. First order stationarity implies that the mean is a constant and second order stationarity implies that the autocovariance at lag i, viz., $E[Q(t_j)\ Q(t_{j+i})]$, where E stands for the expectation, is a function of i only. In streamflow analysis, it is customary to assume weak stationarity, i.e., stationarity with respect to first and second moments.

In general, the first and second moments of flow, viz., mean and variance vary from season to season within an annual cycle. In addition, certain influences of man like the construction of dam may introduce nonstationarity in a flow series. In order to simplify the analysis, it is often the practice to remove the nonstationarity from the time series whenever possible and to analyse the stationary residuals (Roesner and Yevjevich, 1966; Young and Pisano, 1968). For example, the nonstationarity may be due to the presence of long term trends and periodic cycles in the data in which case they can be estimated by standard procedures (Sec.3.3) and eliminated. The residuals then constitute a stationary time series.

2.2 Univariate Models

Mathematical models used to represent a single time series are referred to as univariate models. They may be broadly classified into short lag and long lag models and they are briefly described below.

2.2.1 Short lag models

A p-th order autoregressive (AR) model may be represented as

$$Q(t) = \sum_{i=1}^{p} \phi_i \, Q(t-i) + \varepsilon(t) \tag{2.1}$$

where ϕ_i is the p-th order autoregressive coefficient and $\varepsilon(t)$ is the serially uncorrelated component at time t.

A q-th order moving average (MA) model may be represented as

$$Q(t) = \varepsilon(t) + \sum_{i=1}^{q} \theta_q \, \varepsilon(t-q) \tag{2.2}$$

where θ_q is the q-th order moving average coefficient operating on $\varepsilon(t-q)$, the serially uncorrelated component at time $(t-q)$.

The autoregressive-moving average (ARMA) model constitutes a general class of linear stationary stochastic models and includes the AR and MA models as special cases. A (p, q)

order ARMA model is AR to order p and MA to order q and is described by the equation.

$$Q(t) = \sum_{i=1}^{p} \phi_i \; Q(t-i) + \varepsilon(t) + \sum_{j=1}^{q} \theta_j \; \varepsilon(t-j) \quad (2.3)$$

Eq. (2.3) indicates that the effects of $Q(\cdot)$ and $\varepsilon(.)$ persist for p and q time periods respectively. Generally p and q are equal to or less than 3. Models which have such a short lag are referred to as short lag models. AR, MA and ARMA models have been used extensively in hydrology. For example, first order AR models have been fitted to annual streamflow data (Yevdjevich, 1964) and to monthly streamflow data (Roesner and Yevjevich, 1966). A second order AR model has been fitted to normalised daily streamflow data (Quimpo, 1967). ARMA models of order (1,1) and (2,0) were fitted by Carlson et al., (1970) to annual streamflow data.

The above models are stationary with respect to mean and standard deviation and they also assume covariance stationarity. These assumptions are relaxed and a seasonal variation of the parameters is assumed in a class of univariate models. They may be referred to as seasonally nonstationary univariate models. The Thomas-Fiering model (Maass et al., 1962) was suggested originally for monthly streamflows and is given by.

$$Q(t+1) = \bar{Q}^{j+1} + b^j(Q(t)-\bar{Q}^j) + e(t+1)s^j[1-(r^j)^2]^{\frac{1}{2}} \quad (2.4)$$

where t varies sequentially from, say, 1 to N, the total number of time periods and j varies cyclically from 1 to 12; $\overline{Q}_j$, s_j, r_j and b_j respectively denote the mean flow in the j-th month, standard deviation of flow in the j-th month, correlation coefficient between the flows in the j-th and (j-1)-th months, and the regression coefficient between these two flows. e(t+1) is the serially uncorrelated normally distributed residual at time (t+1) with zero mean and unit standard deviation. As the parameters $\overline{Q}^j$, s^j, r^j and b^j vary with the season j, this is a first order nonstationary AR model.

A more general class of models is one which is nonstationary and whose stationarity can be achieved by repeated differencing of the original series. These are referred to as autoregressive-integrated-moving average (ARIMA) models of order (p, d, q) where d indicates the order of differencing required to render the original series stationary. ARIMA models have been used to describe the monthly streamflow series by McKerchar and Delleur (1972).

2.2.2 Long lag models

Consider the time series (Q(t), t = 1,2,...N). Let the partial sum of flows be given by

$$C(t) = \sum_{i=1}^{t} Q(i) \tag{2.5}$$

The range of cumulative flows R is defined as

$$R = \underset{1 \le t \le N}{\text{Max}} \left[C(t) - \frac{t}{N} C(N)\right] - \underset{1 \le t \le N}{\text{Min}} \left[C(t) - \frac{t}{N} C(N)\right] \tag{2.6}$$

Range is a function of Q(t) and so is itself a stochastic variable. It is one of the indicators of the storage requirements of reservoirs (Fiering, 1967). It can be analytically shown (Feller, 1951) that for pure random and AR series, the N-year range R(N) is proportional to $s^{0.5}$ where s is the standard deviation of the variable. Hurst (1951) empirically analysed natural data and demonstrated that

$$R(N) = 0.61 \; s^H \tag{2.7}$$

where the Hurst coefficient H was found to have an average value of 0.72 in the case of streamflow data. The low lag AR models do not preserve the Hurst coefficient. Use of low lag models then results in an under-representation of extreme events (Mandelbrot, 1968).

Mandelbrot and Wallis (1969) proposed a model for natural processes that exhibit a Hurst coefficient not equal to 0.5. Designated as Fractional Gaussian Noise (FGN) model, this is designed to preserve the Hurst coefficient. In physical terms, FGN has an autocorrelation structure representing the existence of a very long memory where the distant past exerts a small yet significant influence on the present.

Because of the large number of terms and the long lag considered, they are referred to as long lag models. The correlation between successive events for these models may be small and large lag correlations are even smaller and yet their cumulative effect is not negligible. The fast FGN model (Mandelbrot, 1971) is a simplification of the general model. It consists of a low frequency component which is a weighted sum of a number of Markov processes and a high frequency component which is made up of one or more Markov processes. The broken line model (Garcia et al., 1972) also preserves the Hurst coefficient and it has been shown to be an approximation of FGN model.

The application of FGN models to hydrologic data involves the estimation and use of the Hurst coefficient. On the basis of extensive computer experiments, Wallis and Matalas (1970) found that its standard error is quite large and hence for the size of data records normally encountered in streamflow analyses, its estimate is likely to be quite unreliable. The presence of such a long term direct and stationary dependence is rather difficult to justify. Furthermore, ARMA models have been shown to offer a systematic procedure for the representation of complex linear stochastic hydrologic processes and ARIMA models can represent even nonlinear effects. Hence this study is restricted to short lag models only.

2.3 Multivariate Time Series Models

The sequence of values of one variable constitutes a univariate time series. Sequences of values of a number of variables constitute a multivariate time series. The modelling and analysis of multivariate time series should take into consideration not only the serial dependence within each of the time series but also the mutual or crosscorrelation among the time series. Multivariate models in hydrology generally belong to one of the following categories:

2.3.1 Short lag models

In order to reduce the dimensionality of the problem on the basis of dependence among the variables, Fiering (1964) proposed a principal component model for multiple time series. It was shown by Matalas and acknowledged by Fiering (1967) that the above approach does not preserve the correlation structure. Torranin (1972) suggested canonical correlation for application in multivariate hydrologic modelling. Matalas (1967) proposed a multivariate first order AR model given by,

$$\{Q(t+1)\} = [A]\,\{Q(t)\} + [B]\,\{E(t+1)\} \tag{2.8}$$

where $\{Q(t)\}$ is a vector of K variables, say, streamflows at K sites at time t; $\{E(t)\}$ is a vector of K serially and mutually independent variables at time t each with zero mean and unit variance, and [A] and [B] are coefficient matrices.

The components of $\{Q(t)\}$ may also represent different processes at different sites at time t. Young and Pisano (1968) combined the univariate standardisation procedure with the Matalas model and proposed an algorithm for unique estimation of [B] assuming it to have a lower triangular form. A model that lends itself to disaggregation while at the same time preserving the stochastic characteristics at the different levels of disaggregation was formulated by Valencia and Schaake (1972). It consists in first developing a suitable annual flow model which may be stationary. The disaggregation model then develops seasonal flows therefrom. The model due to Beard (1965) consists of multiple regression equations connecting the flow at each site to the previous flows at all sites including itself and current flows at other sites. A set of such multiple regression equations, one for each site, constitutes the multivariate model.

As in the case of short lag univariate models, the short lag multivariate models cannot apparently preserve the Hurst coefficient.

2.3.2 Long lag models.

Long lag models are so designed that they preserve the Hurst coefficient at each site. Matalas and Wallis (1971) have extended the univariate FGN model to two sites. Adopting the Type 2 approximation of FGN, they use the data to find

the mean, variance, skew coefficient, first serial correlation coefficient and Hurst coefficient at each site. Then the crosscorrelation model is set up expressing the crosscorrelation between the flows in terms of the above statistics. The multivariate broken line model (Mejia et al., 1974) uses the lag zero crosscorrelation matrix between the flows at different sites. The lag one crosscorrelation is not preserved. It is an approximation of the FGN model and details have been developed for only a two variate model. Long lag models are reviewed briefly above in order to present the state-of-art in relevent but different approaches to multivariate modelling. They are at present (1976) limited to only two variables and are beset with parameter estimation problems. They are hence not used in this study which is restricted to the consideration of short lag uni- and multivariate models only.

2.4 Decoupled Multivariate Models

2.4.1 Need for a decoupled approach

Let the multivariate vector of streamflows be given by

$$\{Q(t)\} = \{Q_1(t), Q_2(t), \ldots Q_K(t)\}^T$$

where $Q_k(t)$ is the streamflow at site k at time t; K is the total number of sites and T stands for the transpose. To represent the relationship among the time series, a general (p, q) order multivariate ARMA model of the following form

may be considered.

$$\{Q(t+1)\} = [A_1]\{Q(t)\} + \ldots + [A_p]\{Q(t+1-p)\} + [B_1]\{\varepsilon(t+1)\} + \ldots + [B_q]\{\varepsilon(t+2-q)\} \quad (2.9)$$

where $[A_i]$, $i = 1,2,\ldots p$ are the AR coefficient matrices and $[B_j]$, $j = 1,2,\ldots q$ are the MA coefficient matrices, all of dimension $K \times K$, where K is the number of sites considered. This equation can be considered as an extension of the univariate ARMA model to the multivariate domain. With adequate data, this equation can be solved for the unknowns. But because of the large number of parameters, the results are likely to be data dependent and unstable.

The complication in multivariate modelling is due to the fact that the variability of multisite streamflows is caused both by the internal correlation of flows at each site and the external correlation among the flows at the various sites. It seems possible and desirable to decouple the two and deal with them separately. A univariate stochastic model may first be fitted to each of the flow series thus accounting for the internal correlation. The resulting residuals are referred to as prewhitened series since they are serially independent and normally distributed. A multivariate model can be fitted to the univariate residual series in terms of serially and mutually uncorrelated random components to explain external correlations.

Such an approach for the modelling of multivariate time series seems to be advantageous because (i) it enables the utilisation of techniques of univariate time series analysis, which are quite advanced; (ii) as the parameters of the univariate and multivariate models are estimated separately, their number is comparable to that of univariate and multivariate models respectively and the question of simultaneous estimation of all the parameters is avoided, and (iii) it affords flexibility in the choice of appropriate but different models for the univariate and multivariate time series. Hence the decoupling approach seems to hold promise of wide application in multivariate modelling of interrelated time series.

2.4.2 Other decoupled models

Kareliotis and Chow (1972) suggested a procedure for the analysis of residual hydrologic stochastic processes which may be considered as a special case of the decoupled model.

Yevjevich and Karplus (1972) have introduced the decoupling approach by first determining the univariate residuals at each site and connecting these residuals by distance-correlation equations. Frost and Clarke (1973) fitted first order AR models to the streamflow at two stations, and a multivariate white noise model, to the residuals. Their model is equivalent to assuming a diagonal form for the [A] matrix in Eq. (2.8) and estimating the corresponding [B] matrix. Recently Yevjevich (1975) has fitted second order AR models to the net basin supplies of four of the Great Lakes,

and a multivariate white noise model to the univariate residuals. The decoupled multivariate models developed in these studies are stationary models, in which the correlation structure is assumed to be invariant with respect to time.

2.4.3 Methodology proposed for the decoupled model

The mathematical modelling of multisite streamflow sequences by the proposed decoupled multivariate model involves the following steps: i) normalisation of streamflows at each station by logarithmic, square root or other transformations as appropriate; ii) standardisation of streamflows at each site to achieve stationarity in mean and standard deviation; and iii) fitting a univariate model at each site to separate the persistence and serially independent components. It is proposed to use only short lag models in this study. Suitable models include those of Box and Jenkins (1970), Thomas-Fiering (Maass et al., 1962) etc. For example, the univariate ARMA model is given by,

$$x(t+1) = \phi_1 x(t) + \phi_2 x(t-1) + \ldots + \phi_p x(t+1-p) + \theta_1 \varepsilon(t+1) + \theta_2 \varepsilon(t) + \ldots + \theta_q \varepsilon(t+2-q) \qquad (2.10)$$

where $x(t)$ is the normalised, standardised flow at time t, $\varepsilon(t)$ is the serially independent component at time t and ϕ_i and θ_j are respectively the i-th order AR and j-th order MA

coefficients. This may be extended for a number of sites as,

$$\{X(t+1)\} = [\phi_1]\{X(t)\} + \ldots + [\phi_p]\{X(t+1-p)\}$$
$$+ [\theta_1]\{\varepsilon(t+1)\} + \ldots + [\theta_q]\{\varepsilon(t+2-q)\} \quad (2.11)$$

where $\{X(t)\}$ is the vector of normalised standardised streamflow series at time t, $\{\varepsilon(t)\}$ is the vector of the univariate residuals at time t and $[\phi_i]$ and $[\theta_j]$ are diagonal matrices respectively representing the i-th order AR and j-th order MA coefficients of the univariate models at the several sites under study. In Eqs. (2.10) and (2.11), the univariate models are assumed to be stationary. The model can also be modified appropriately to account for nonstationarity when seasonal variations in the correlation structure are taken into account. After the univariate models are fitted, the serially independent random components at each site may be tested for their randomness and distribution; iv) Let $\{E(t)\}$ be the vector of univariate residuals obtained by standardising the $\{\varepsilon(t)\}$ vector viz.,

$$\{E(t)\} = \{\varepsilon(t)/s_\varepsilon^j\} \quad (2.12)$$

where s_ε^j is the standard deviations of univariate residuals for season j. Let $\{\eta(t)\}$ be a vector of pure random components at time t which are free from serial and crosscorrelation. Then $\{E(t)\}$ may be related to $\{\eta(t)\}$ by a (p,q) order multivariate ARMA model. In this study, the simpler Matalas model corresponding to p = 1 and q = 1 was found to be satisfactory, viz.,

$$\{E(t+1)\} = [C]\,\{E(t)\} + [D]\,\{\eta(t+1)\} \tag{2.13}$$

where [C] and [D] are the coefficient matrices. Depending on the nature of streamflow data, [C] and [D] may be constant coefficient matrices that do not change with time or season and hence are stationary; or because of seasonal variations in crosscorrelation structure, they may vary with season or time and hence are nonstationary; v) By reversing the steps, the coupled multivariate model can be derived. It may be shown from Eqs. (2.11) and (2.12), that the coupled relationship is of the form,

$$\begin{aligned}\{X(t+1)\} &= [\phi_1]\,\{X(t)\} + \ldots + [\phi_p]\,\{X(t+1-p)\} \\ &+ [\theta_1]\,[s_\varepsilon^{j+1}]\,[C]\,\{E(t)\} + \ldots + \\ &+ [\theta_q]\,[s_\varepsilon^{j-q+2}]\,[C]\,\{E(t+1-q)\} \\ &+ [\theta_1]\,[s_\varepsilon^{j+1}]\,[D]\,\{\eta(t+1)\} + \ldots \\ &+ [\theta_q]\,[s_\varepsilon^{j-q+2}]\,[D]\,\{\eta(t+2-q)\}\end{aligned} \tag{2.14}$$

In the right hand side of Eq. (2.14), the first p terms represents the AR component with reference to the earlier values of the same time series, the next q terms represent the crosscorrelated moving average terms among the serially independent random components and the last q terms represent the crosscorrelation among the pure random components. For the case of stationary univariate first order AR models at all the sites, p = 1, q = 1 and $[\theta_1] = [I]$, so that Eq. (2.14)

becomes,

$$\{X(t+1)\} = [\emptyset_1]\{X(t)\} + [s_\varepsilon^{j+1}][C]\{E(t)\} + [s_\varepsilon^{j+1}][D]\{\eta(t+1)\} \quad (2.15)$$

Knowing the mean and standard deviation of the transformed flows $\{\bar{Y}^j\}$ and $\{s_y^j\}$ for the season j, $\{X(t+1)\}$ may be destandardised with $[s_y^{j+1}]$ diagonal matrix with elements s_y^{j+1}.

$$\{Y(t+1)\} = \{\bar{Y}^{j+1}\} + [s_y^{j+1}][\emptyset_1]\{X(t)\} + [s_y^{j+1}][s_\varepsilon^{j+1}][C]\{E(t)\} + [s_y^{j+1}][s_\varepsilon^{j+1}][D]\{\eta(t+1)\} \quad (2.16)$$

Depending upon the normalising transformation used, the $\{Y(t)\}$ series can be transformed back to the original $\{Q(t)\}$ series.

The proposed decoupled multivariate model was developed independently of Frost and Clarke (1973) and Yevjevich (1975) and differs from the earlier ones in a number of features, viz.,

i) The decoupling of the within-the-station correlation and among-the-station correlation as well as prewhitening of the univariate series are specifically identified;

ii) Different univariate and multivariate models may be fitted to the time series as appropriate;

iii) The recoupled equations are derived in terms of the decoupled models and

iv) The model is formulated in terms of general non-stationary univariate and multivariate models to facilitate incorporation of seasonal variation of within-the-station and among-the-station correlations.

3. PRELIMINARY ANALYSIS OF HYDROLOGIC DATA

3.1 Data Used for the Study

Streamflow records for three major tributaries to a river in North India were available for the study. The average daily flow was expressed in cusecs and there were three values per month corresponding to the first ten days, second ten days and the remainder of the month. These are referred to as ten daily flows. Monthly average flows were estimated as the average of the three consecutive ten daily flow values. Similarly average annual flows were estimated as the average of the twelve monthly values. The period over which concurrent data were available was 25 years only. Even though records were available in some stations for a longer period, only concurrent data were used in this study.

3.2 Frequency Analysis of Streamflow Data

3.2.1 Normalisation of data

Normal distribution for the random variables confers some mathematical advantages. It is well analysed and documented. Also, by virtue of its linearity, viz., that a linear combination of normally distributed variables yields a normal

variable, it is amenable to algebraic manipulation. For instance, bivariate regression and correlation models such as the Thomas-Fiering model(Maass et al.,1962) are based generally on the assumption of normality. Another special feature of the normal distribution is with regard to the stationarity of the underlying stochastic process. Strict stationarity requires that the moments of all orders remain time invariant. But the estimation of moments of high orders is unreliable especially if the sample is short. Hence satisfying the requirement of stationarity becomes difficult. But if the underlying distribution is normal, stationarity of the first two moments implies strict stationarity. Estimation of the parameters from these two moments can be expected to be reliable.

Streamflow data can have only nonnegative values and further the resulting distribution is generally skewed. In such cases, a skewed distribution may have to be fitted to the data. In order that the advantages of the normal distribution may still be utilised, the skewed natural flow series may be transformed to a normal distribution. This is the basis of normalisation of data. The procedures generally adopted for normalisation include (i) N-th root transformation (Stidd, 1970), (ii) logarithmic transformation (Beard, 1965; Young et al., 1968) and (iii) Hilferty-Wilson transformation (Beard, 1965) for gamma distributed data.

(i) N-th root transformation involves the derivation of the N-th root of the observations so that the transformed series follows a normal distribution. Square and cube root transformations have been found to help in normalisation. They were not found to be valid for the data used in this study.

(ii) Logarithmic transfomation (Chow, 1964) involves the consideration of the natural or common logarithm of the variable. When the lower bound of the variable is nonzero, the probability distribution is referred to as a three parameter logarithmic distribution. In case one or more observations lie at the lower bound, it may be necessary to add a constant to all the observations in order that the logarithms are finite. For the data used in this study, the additive term was not found necessary. Logarithmic transformation has been found to be very useful in transforming highly skewed and nonnegative variables like streamflows to nearly normally distributed ones.

(iii) If the original distribution is a gamma distribution (Pearson Type 3 distribution), it can be transformed to a normal distribution by the Hilferty-Wilson transformation (Beard, 1965). The procedure consists in first standardising the raw series so that the resulting series has zero mean and unit standard deviation. Let x_i be the Pearson Type 3 standard deviate obtained from the standardised data with

skew coefficient g. Then the corresponding normal (0, 1) deviate t_i can be obtained by the transformation

$$t_i = \frac{6}{g}\left[\left(\frac{gx_i}{2}+1\right)^{1/3} - 1\right] + \frac{g}{6} \qquad (3.1)$$

In case the probability distribution is a log Pearson Type 3 distribution, it is necessary to use the logarithmic transformation, standardisation and then Hilferty-Wilson transformation.

3.2.2 Fitting a probability distribution

Probability distributions may be fitted to random variables on the basis of available observations. This involves the assumption of a suitable theoretical model, estimation of parameters and testing the goodness of fit.

Estimation of parameters: The parameters may be estimated generally by one of the three following methods (Chow, 1964; Yevjevich, 1972):

i) Method of moments (MM): Method of moments relates the sample values of the moments to the parameters of the distribution. The r-th sample moment about any arbitrary x(0) is given by,

$$m_r = \frac{1}{N} \sum_{i=1}^{N} [x(i) - x(0)]^r \qquad (3.2)$$

where N is the size of the sample. For any distribution, the first moment about the origin gives the sample mean and the second moment about the mean gives the sample variance, viz.,

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x(i) \tag{3.3}$$

and

$$\mathrm{Var}\ (x) = \frac{1}{(N-1)} \sum_{i=1}^{N} (x(i) - \bar{x})^2 \tag{3.4}$$

$\bar{x}$ and Var (x) are the sample estimates of μ and σ^2 for the theoretical distribution.

Estimation by the method of moments is asymptotically efficient and the efficiency is usually smaller than unity. For small samples, the estimates are significantly affected by extreme or off-control points that may be present in the sample.

Let μ and σ be the mean and standard deviation of the lognormally distributed variable. Using the MM, the mean $\mu(n)$ and the standard deviation $\sigma(n)$ of the normalised variable are given by

$$\mu(n) = \frac{1}{2} \ln \left(\frac{\mu^4}{\mu^2 + \sigma^2}\right) \tag{3.5}$$

and

$$\sigma(n) = \left[\ln \left(\frac{(\sigma^2 + \mu^2)}{\mu^2} \right)\right]^{\frac{1}{2}} \tag{3.6}$$

ii) Method of least squares (MLS): This method consists in the estimation of the parameters of the assumed distribution by minimizing the sum of squares of deviations of the observed points from the fitted function. Chow (1964) suggests the following procedure: If $\bar{x}$ and s_x are the sample mean and standard deviation, the frequency factor K is defined as

$$K = \frac{x - \bar{x}}{s_x} \tag{3.7}$$

For any given distribution, K can be related to the cumulative distribution function F(x) of the distribution. For a given x_i, $F(x_i)$ is estimated from plotting position formulae. For the assumed distribution, K_i is obtained from $F(x_i)$. The regression line of x on K is determined by MLS. This gives estimates of sample mean $\bar{x}$ and standard deviation s_x. Though this procedure is not theoretically exact, it generally gives a better overall fit than the MM, and further, the estimate is not affected as much by extremely rare occurrences as in the case of MM.

iii) Method of maximum likelihood (MLM): Maximum likelihood estimate is that estimate of the parameters of a distribution for which the probability of occurrence of the actual observations is a maximum.

Let x(1), x(2), ... x(N) be the N observations and f(x), the assumed probability density function in terms of the

parameters p(1), p(2), ... p(J). Assuming independence of events, the probability of the outcome P is given by

$$P = \prod_{i=1}^{N} f[x(i)] \tag{3.8}$$

For this to be a maximum, $\partial P/\partial p_j = 0$ for all j = 1,2,...J. The J equations in terms of the J parameters are solved to give the maximum likelihood estimates. For the normal distribution, these estimates are the same as the estimates by the MM. For the two parameter lognormal distribution, the MLE of mean and standard deviation are given by the mean and standard deviation of the logarithm of the sample.

Test for the goodness of fit: In the present study the goodness of fit of the fitted distribution to the observed data is tested by the chisquare test. The sample space is divided into I mutually exclusive classes with a class frequency of 5 or more. Let p(i) be the probability that the variable belongs to the i-th class for the assumed distribution. If x(i) and x(i+1) are the limits of the i-th class interval, then

$$p(i) = F[x(i+1)] - F[x(i)] \tag{3.9}$$

Let f(i) be the observed frequency of the sample from the i-th group. If N is the total number of samples, the chisquare statistic is given by

$$\chi^2 = \sum_{i=1}^{I} \frac{(f(i) - N(p(i))^2}{Np(i)} \qquad (3.10)$$

If J is the number of parameters estimated, then theoretically χ^2 has a chisquare distribution with (I - J - 1) degrees of freedom. Let $\chi^2(\alpha)$ denote the value of χ^2 at $\alpha\%$ confidence level for the above degrees of freedom as obtained from the tables. If the calculated χ^2 is greater than the theoretical value, then the sample deviates significantly from the assumed distribution at the given level and the fit is rejected. If it is less, then the fit is accepted.

In the present study, the transformed standardised data were tested for normality. The range of the distribution was generally (viz., for all except in the case of the annual series) divided into six classes. They are respectively $-\infty$ to -1.0, -1.0 to -0.5, -0.5 to 0.0, 0.0 to 0.5, 0.5 to 1.0 and 1.0 to ∞. Probability distribution function values were obtained from statistical tables and they were used to estimate the probability values for the class intervals. A value of p = 95% was chosen as the confidence level.

Annual discharges: The mean and standard deviation of flows were estimated by MM (Table 3.1). For all the three rivers, the hypothesis of normal distribution was accepted at 95% level.

TABLE 3.1 χ^2 ESTIMATES OF ANNUAL FLOWS (MM) FOR NORMAL DISTRIBUTION

RIVER	1	2	3
χ^2 (ESTIMATED)	5.3	2.1	10.0
DEGREES OF FREEDOM	10	4	9
χ^2 (THEORETICAL) AT 95% LEVEL	18.3	9.5	16.9

Monthly discharges: The streamflow data corresponding to each month were analysed for frequency distribution using parameter estimates by MM. The untransformed data were tested for normality (Table 3.2). Several cases were found to be

TABLE 3.2 χ^2 ESTIMATES OF MONTHLY FLOWS (MM) FOR NORMAL DISTRIBUTION.
(DEGREES OF FREEDOM = 3, χ^2_{95} = 7.8)

MONTH / RIVER	1	2	3	4	5	6	7	8	9	10	11	12
1	1.3	0.3	3.5	11.6+	18.1+	0.8	12.0+	7.1	8.7+	3.3	2.1	1.0
2	1.7	1.3	3.3	1.3	9.9+	3.1	0.7	0.7	8.1+	1.9	0.8	8.8+
3	1.0	2.6	1.6	3.9	9.9+	9.9+	1.5	3.4	0.6	4.0	3.2	13.3+

+ Indicates significant values

significantly different from normal distribution at 95 % level. Hence the hypothesis of normality of the original data was rejected. Several procedures for normalisation including the

N-th root transformation, Hilfrety-Wilson transformation and logarithmic transformation were tried. Based on results, logarithmic transformation was adopted in this study. Standardisation (Subsec. 3.3) was done using parameters estimated respectively by MM, MLS using Chow's frequency factor and MLE (Table 3.3). The goodness of fit was tested in each case. Only one value by MM was found to be significant at 95% level for river 1; all values were insignificant for river 2 and for river 3 MLS had 3 significant values and MLE 2 significant values. However all except one of the above values were insignificant at 95% level. Hence lognormality was assumed to be valid whatever be the method used for estimation of parameters.

Tendaily discharges: Streamflow data were grouped into 36 tendaily seasonal values. Using parameter estimates by MM, chisquare test was performed to test normality. The fit was found to be not good at 95% level for 15, 6 and 11 periods respectively for rivers 1, 2 and 3, and so the assumption of normality was rejected. The data were then logtransformed using parameters estimated respectively by MM, MLE and MLS and chisquare test was applied. The results are shown in Table 3.4. It is seen that logtransformation improves the fit significantly and the 3 methods of standardisation are comparable though individual variations exist from river to river. The fit of log-Pearson Type 3 distribution was also

TABLE 3.3 χ^2 ESTIMATES OF LOGTRANSFORMED MONTHLY SERIES FOR NORMAL DISTRIBUTION
(DEGREES OF FREEDOM = 3, χ^2_{95} = 7.8)

MONTH / METHOD	1	2	3	4	5	6	7	8	9	10	11	12
					RIVER 1							
MM	0.7	1.0	0.6	4.9	9.2+	0.2	3.2	5.8	1.2	1.0	4.5	2.5
MLE	0.7	1.5	0.6	4.0	6.2	0.2	2.6	3.5	1.2	1.0	4.5	1.3
MLS	0.2	2.0	1.1	1.4	4.4	0.3	3.1	5.8	1.2	0.9	5.9	1.3
					RIVER 2							
MM	0.7	2.1	1.3	1.5	5.7	3.5	0.7	0.7	7.7	0.6	0.8	0.9
MLE	0.7	2.1	2.5	2.6	1.2	3.5	0.7	0.2	2.2	0.6	0.8	1.6
MLS	1.2	0.7	2.5	0.7	1.3	1.7	0.8	0.2	4.4	0.7	1.4	1.4
					RIVER 3							
MM	0.8	1.6	0.6	0.7	5.7	2.3	2.0	5.0	0.7	2.7	2.0	7.3
MLE	0.4	1.6	0.8	0.3	5.6	2.2	1.5	8.3+	0.7	2.7	1.0	8.6+
MLS	0.4	3.0	0.8	0.3	8.4+	2.8	2.0	8.3+	0.8	2.7	2.7	8.6+

+ Indicates significant values

TABLE 3.4 RESULTS OF χ^2 TEST
DATA

DISTRIBUTION	RIVER	Number of Cases Significant at 95% Level					
		Monthly Series			Ten daily Series		
		1	2	3	1	2	3
NORMAL	(MM)	4	3	3	15	6	11
LOGNORMAL	(MM)	1	0	0	3	3	3
	(MLE)	0	0	2	5	1	3
	(MLS)	0	0	3	2	0	4
LOG-PEARSON TYPE 3					3	1	5

tried by applying the Hilferty-Wilson transformation to logtransformed data standardised with the parameters of MLE. There seems to be no significant improvement because of Hilferty-Wilson transformation. Hence only logarithmic transformation was adopted in this study.

3.3 Standardisation of Data

The parameters of the process, viz., mean and standard deviation of the monthly and tendaily data exhibit a change with season. This is referred to as cyclic nonstationarity, which has to be removed in order to make the process stationary.

Standardisation refers to the conversion of the above series to a series with zero mean and unit standard deviation. Two procedures are generally used for this purpose:

i) Nonparametric standardisation: Nonparametric standardisation consists in estimating the seasonal mean and standard deviation $\bar{y}^j$ and s_y^j and using them to transform the given series y(i) to a series x(i) with zero mean and unit standard deviation, as follows:

$$x(i) = \frac{y(i) - \bar{y}^j}{s_y^j} \tag{3.11}$$

ii) Parametric standardisation: Yevjevich (1971) suggests the fitting of a harmonic series each for the mean $\bar{y}^j$ and standard deviation s_y^j. Let the harmonic function fitted for the mean and standard deviation be $\widetilde{y}^j$ and $\widetilde{s}_y^j$ respectively. Let

$$Z(i) = \frac{y(i) - \widetilde{y}^j}{\widetilde{s}_y^j} \tag{3.12}$$

The Z(i) series will not have exactly zero mean and unit standard deviation. It is once again standardised to another series using the mean $\bar{Z}$ and standard deviation s_Z of the Z(i) series. The new series is stationary in mean and standard deviation. Generally this procedure requires the estimation of less number of parameters than nonparametric standardisation. In this study, the earlier and more general procedure of nonparametric

standardisation was followed to obtain standardised data. Standardisation was carried out separately for monthly and tendaily data. using the parameters estimated by each one of the three methods, viz., MM, MLE and MLS. The normalisation and standardisation of streamflow data constitute the preliminary analyses. The normalised standardised data are used subsequently in univariate and multivariate time series analyses.

4. UNIVARIATE STOCHASTIC MODELS

4.1 ARIMA Models

A general introduction to univariate models has already been given in Sec. 2.1. Univariate modelling as adopted in this study is presented in this chapter. The class of nonlinear models known as ARIMA models are often used for the representation of univariate time series. If the time series is nonstationary, differencing of the series is done a number of times if necessary till stationarity is achieved. A (p, d, q) ARIMA model is one where the d-th differencing results in a model which is autoregressive to order p and moving average to order q, viz., it can be represented by

$$\nabla^d x(t) = \sum_{i=1}^{p} \phi_i \; x(t-i) + \varepsilon(t) + \sum_{j=1}^{q} \varepsilon(t-j) \quad (4.1)$$

where ∇^d refers to the d-th order differencing; the first group of terms in the right hand side refers to the AR component; and the second group refers to the moving average component. A special case of the ARIMA model is the ARMA (p, q) model where no differencing is needed. In selecting the ARIMA model for univariate analysis, the following advantages were noted:

(i) availability of a general methodology of selection among alternatives, and (ii) availability of an efficient computational procedure which will locate the parameters of the ARIMA model (Box et al., 1970, McKerchar et al., 1972, Nelson, 1973). The procedure for ARIMA and ARMA models are the same except that in the former the first step is to determine the order of differencing d. No differencing was needed for the normalised standardised data series and hence the methods for ARMA model alone are briefly dealt with in the following section.

4.1.1 Identification of the model

Identification refers to the choice of a particular model from the general family that is most appropriate to the sample time series and is helped by determining the correlation structure. The theoretical autocorrelation coefficient ρ_i is the ratio of the autocovariance at lag i to the variance of the process. In the present study, the time series has already been standardised. So the sample autocorrelation r_i is equal to the sample autocovariance and is given by

$$r_i = \frac{1}{N-i} \sum_{j=1}^{N-i} x_j \, x_{j+i} \qquad (4.2)$$

where N is the number of observations and x_j is the value of the random variable at time j. Given the sample data, an

estimate of the autocorrelation function may be obtained by computing r_i with $i = 1, 2, \ldots$. The largest possible subscript i would be N-1 although the computation is often carried to considerably less than N-1 lags. A visual device which is very useful in identification is the sample correlogram, a graph of sample autocorrelations. The object in studying the sample autocorrelation is to recognize in them a pattern typical of an ARMA process whose correlogram structure is familiar.

It is to be noted that sample autocorrelations are only estimates of the actual autocorrelations for the process which has generated the data on hand. The sample autocorrelations are hence subject to sampling errors, particularly for the small sizes of data usually available for streamflows. Consequently sample autocorrelations may sometimes be largely in error and so they may not be indicative of the true autocorrelations. Hence while interpreting the sample correlogram, the general characteristics which are recognizable in the sample correlogram should only be looked into and smaller details ignored. To distinguish what is significant from what is not, a test for statistical significance of sample autocorrelations may be made. For a random normal series, the autocorrelations should be normally distributed about zero with variance $\frac{1}{N}$, so that all autocorrelations,

beyond $\pm$ 1.96 times the standard errors may be considered significant at 95% level (Yevjevich, 1972).

For a moving average process, the correlogram exhibits a cutoff at a certain lag beyond which it is not significant, thus facilitating the determination of the order of the moving average process. For autoregressive models, the correlogram decays slowly and is not helpful in indicating the order of the model. In such cases a set of sample statistics, known as partial correlation coefficients helps in the estimation of the orders of autoregression.

Consider a pure autoregressive model of order p, viz.,

$$x(t) = \sum_{i=1}^{p} \phi_i \, x(t-i) + \varepsilon(t) \tag{4.3}$$

where ϕ_i is the AR coefficient of order i and $\varepsilon(t)$ is the residual. These coefficients are related to the autocorrelation coefficients by the Yule-Walker equations given by

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \cdot \\ \rho_p \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \cdots & \rho_{p-2} \\ \cdots & \cdots & \cdots & \cdots \\ \rho_{p-1} & \rho_{p-2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \cdots \\ \phi_p \end{bmatrix} \tag{4.4}$$

If ρ_1, ρ_2, ... ρ_p and p are known the above system of p equations with p unknowns ϕ_1, ϕ_2, ..., ϕ_p can be solved. In practice, the true values of p and ρ_i are unknown. Let p = 1. Using Yule-Walker equations and using the sample

value r_1 for ρ_1, the population value, one gets $r_1 = \hat{\phi}_1$ where $\hat{\phi}_1$ is the resulting estimate of ϕ_1. If $\hat{\phi}_1$ is significantly different from zero, it can be concluded that the process is at least of order one. To see whether the process is of order two or greater, the Yule-Walker equations are solved for $p = 2$,

$$
\begin{aligned}
r_1 &= \hat{\phi}_1 + \hat{\phi}_2 r_1 \\
r_2 &= \hat{\phi}_1 r_1 + \hat{\phi}_2
\end{aligned}
\tag{4.5}
$$

If the resulting estimate of ϕ_2 differs significantly from zero, it can be concluded that the process is atleast of order two. This procedure is repeated successively for larger values of p. If the true order of the model is p_t, then, when the system of equations is solved for $p = p_t + 1$, the value of $\hat{\phi}_{p_t+1}$ will not be significantly different from zero since it is an estimate of ϕ_{p_t+1} which is zero. Denoting by $\hat{\phi}_{ii}$ the value of ϕ_i implied by the solution for $p = i$, the $\hat{\phi}_{ii}$ are referred to as the estimated partial correlation coefficients of the process. If the order of autoregression is p_t, then

$$
\hat{\phi}_{ii} = 0 \quad \text{for } i > p_t \tag{4.6}
$$

Since the partial correlation coefficients are sample statistics and therefore subject to sampling error, a test is needed to decide when $\hat{\phi}_{ii}$ is indistinguishable from zero in a

statistical sense. Under the hypothesis that $p = p_t$, the approximate standard error of $\hat{\phi}_{ii}$ is given by

$$SE\ (\hat{\phi}_{ii}) \quad = \quad \frac{1}{\sqrt{N}} \quad \text{for } i > p \tag{4.7}$$

Thus it is inferred that $p = p_t$ if $\hat{\phi}_{p_t+1,\ p_t+1}$ is small compared to $1.96/\sqrt{N}$. Hence the partial correlation function does for the AR model what the correlogram does for the moving average model. If the process is pure moving average, then the partial correlation coefficients decline in magnitude with increasing lag and do not exhibit cutoff at any lag. A mixed ARMA model is identified by a gradual decline in both the correlogram and the partial correlation coefficients.

In order to test whether the flow series is a realisation of a serially uncorrelated process, one way is to see if the correlogram ordinates are within roughly twice the standard error (i.e., at 95 confidence level). But the standard error underestimates the standard deviation of sample autocorrelation especially at low lags. To obviate this difficulty, a 'Q' statistic that offers a test on the smallness of a whole set of sample autocorrelations for lags 1 to K has been suggested (Box et al., 1970) viz.,

$$Q \quad = \quad N \sum_{i=1}^{K} {r_i}^2 \tag{4.8}$$

The Q statistic is approximately chisquare distributed with (K-p-q) degrees of freedom.

Annual series: The annual data were used to estimate the correlogram and partial correlogram shown in Figs. 4.1 and 4.2. They indicate that AR or MA processes are not present and the series may be considered random. The partial correlogram had significant values at lags 8 and 10 for river 2 and only at lag 10 for river 1. The intermediate orders have insignificant coefficients. As the sample size is only 25, the estimates at higher lag may not be reliable. The Q statistics in the three cases are 8.8, 12.4 and 10.4, and are less than the theoretical chisquare value of 18.31 at 10 degrees of freedom at 95 percent confidence level. Hence the hypothesis of independence of annual flows is accepted and the annual flow series may be considered as serially independent.

Monthly series: Correlograms of the standardised monthly series of the 3 rivers are shown in Fig. 4.3. They show a gradually declining trend indicating an AR process. The partial autocorrelation coefficients (Fig. 4.4) exhibit a significant value at lag 1 and insignificant values at higher lags suggesting a first order AR model. Further, the serial independence of the monthly flows was tested using the Q statistic. The statistic had values far higher than the permissible value at 95 % level indicating the rejection of the hypothesis of serial independence. So a first order AR model seems to be suitable for logtransformed monthly flows.

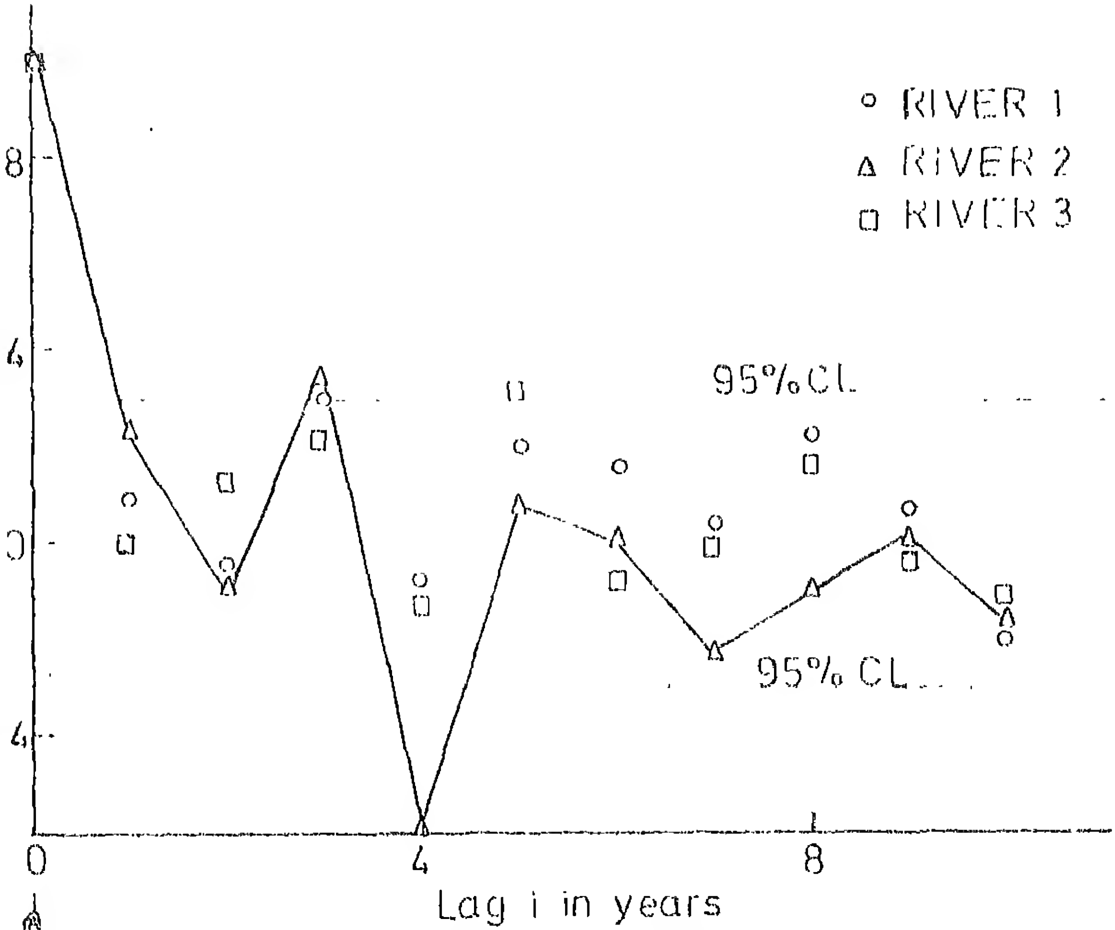
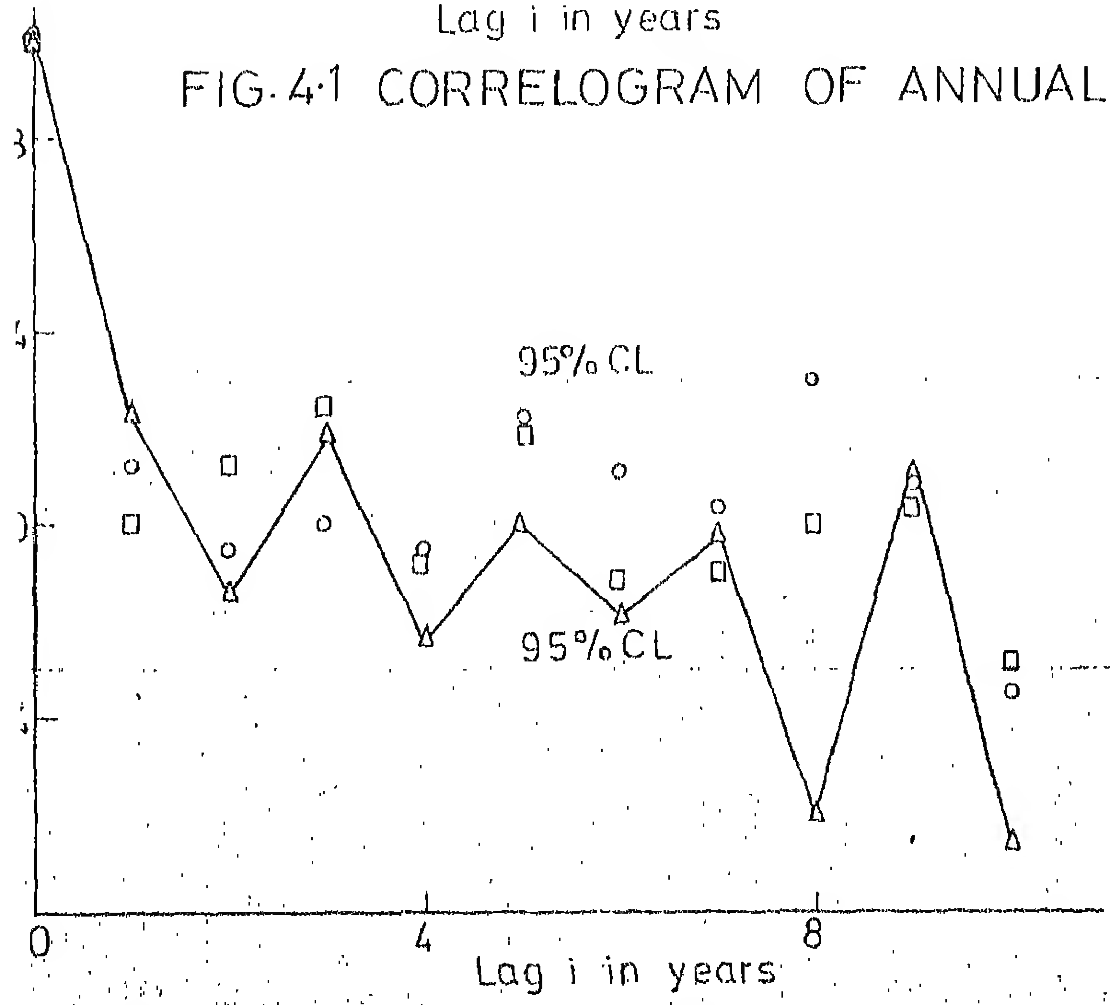


FIG. 4.1 CORRELOGRAM OF ANNUAL FLOWS



4.2 PARTIAL CORRELOGRAM OF ANNUAL FLOWS

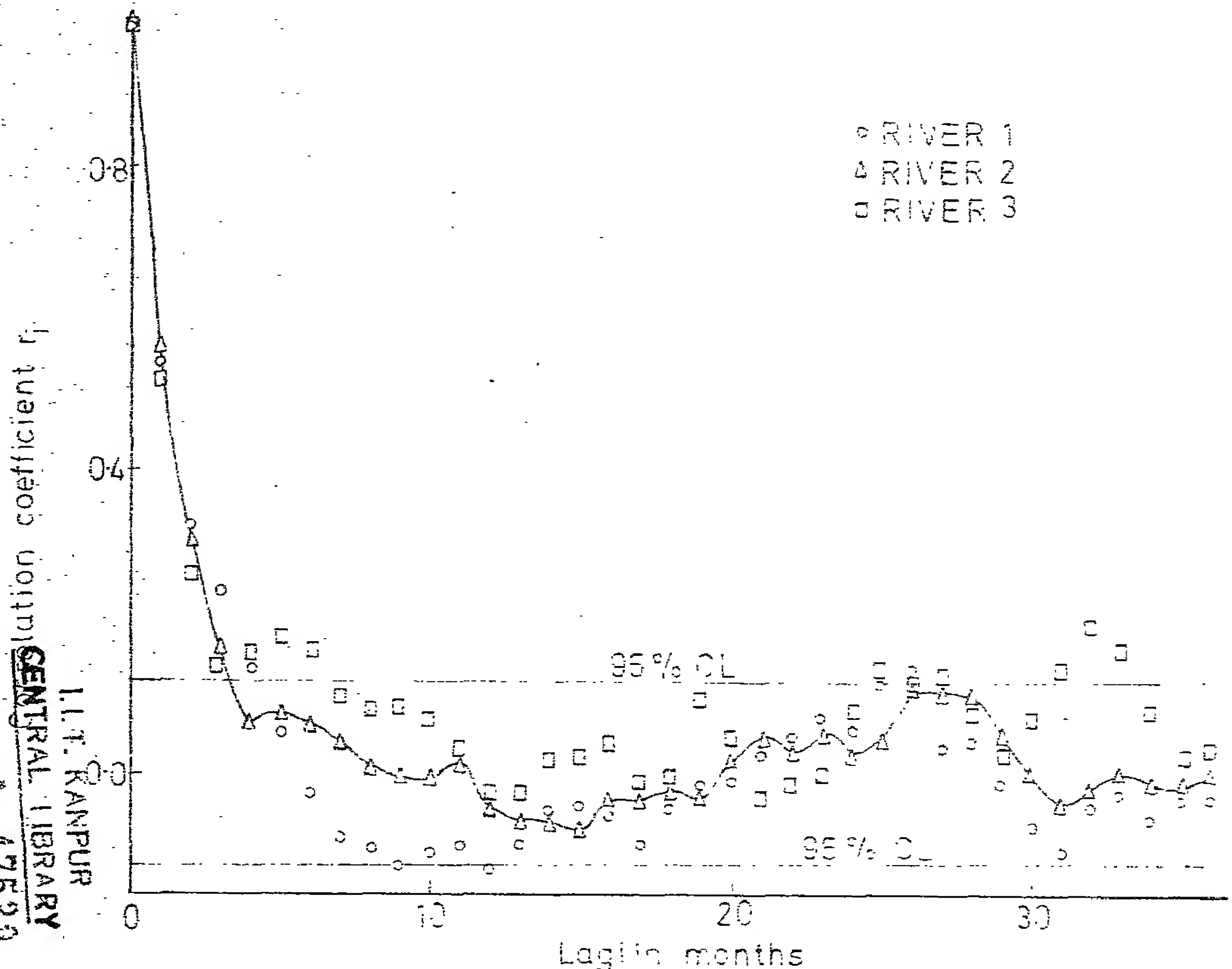


FIG. 4.3 CORRELOGRAM OF NORMALISED STANDAPDISED MONTHLY SERIES

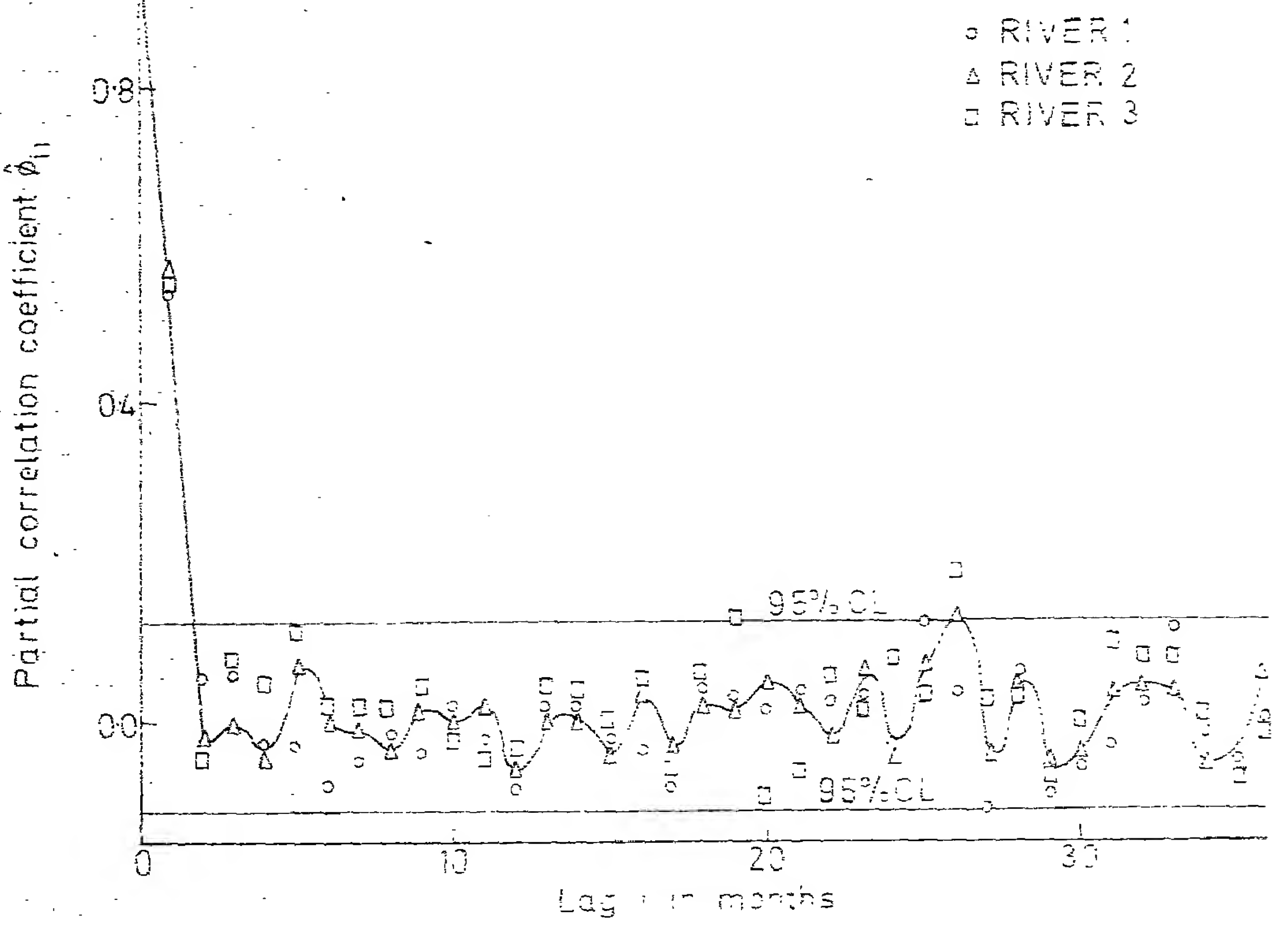


FIG. 4·4 PARTIAL CORRELOGRAM OF NORMALISED STANDARDISED MONTHLY SERIES

Tendaily series: The correlograms of the standardised tendaily series shown in Fig. 4.5 exhibited a much slower decay than the monthly series and the partial correlation coefficients (Fig. 4.6) had significant values at 95% level at lags 1 and 3 for rivers 1,2 and 3; and lag 2 also for rivers 1 and 3. The value for river 2 at lag 2 is nearly significant at 95% level. The presence of monthly persistence was already borne out by a significant lag one coefficient in the monthly time series. The Q statistic for the three rivers were respectively 1110, 1078 and 1080 at 10 degrees of freedom, being very much greater than the theoretical value of 18.31. Hence the hypothesis of serial independence was rejected. A third order AR model is hence indicated for the tendaily series. It may be possible to fit as an approximation a (1,1) ARMA model from parsimony considerations. But this was not adopted in this study.

4.1.2 Estimation of parameters

For a pure AR process, the estimation of the starting values of the ϕ coefficients is straightforward and is done using the Yule-Walker equations relating these coefficients to the ordinates of the correlogram. The solution of this system of equations supplies the starting estimates.

The relationship between autocorrelation coefficients and parameters is nonlinear for moving average and mixed processes and hence the computation of initial estimates is more involved. Generally the following procedure is

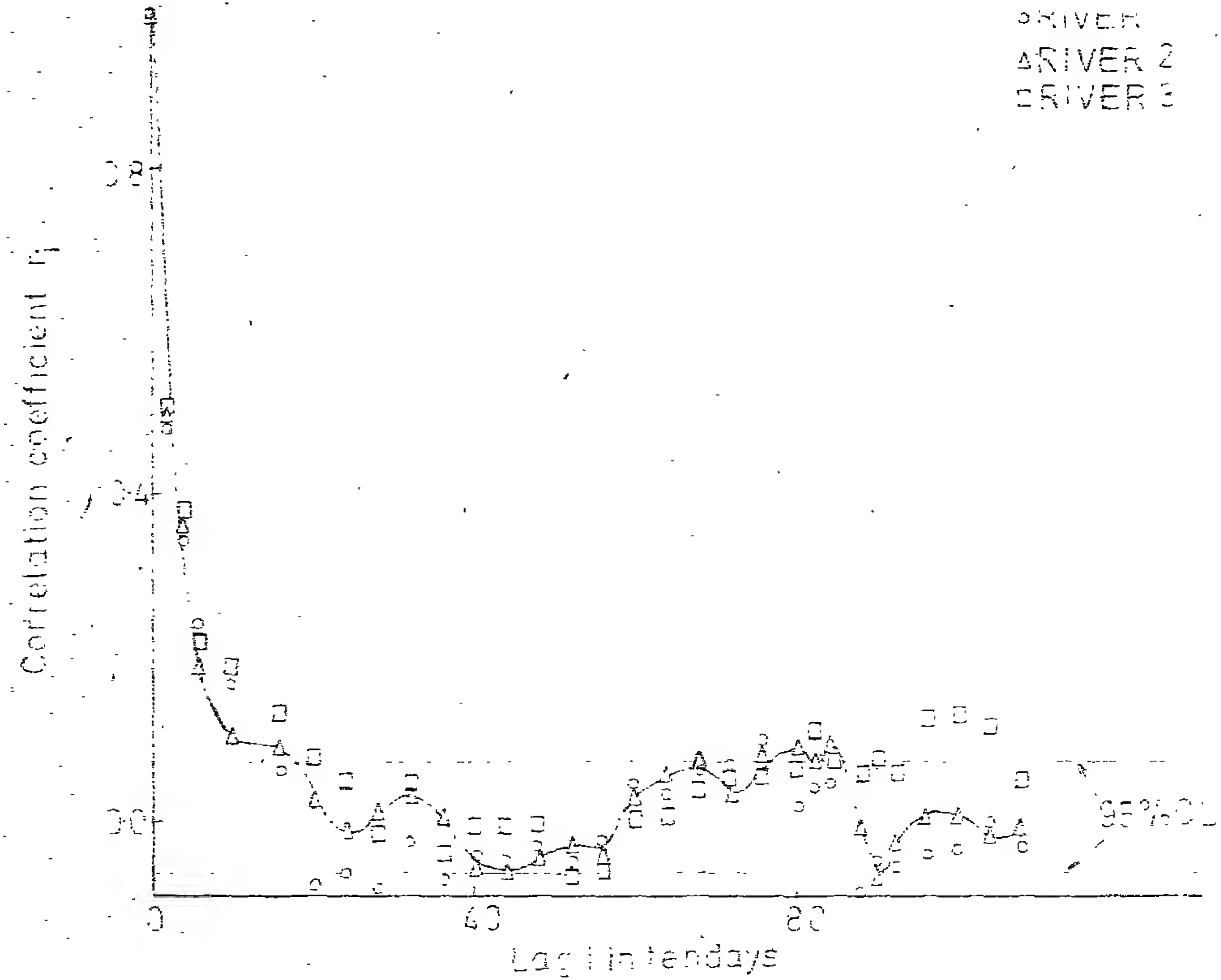
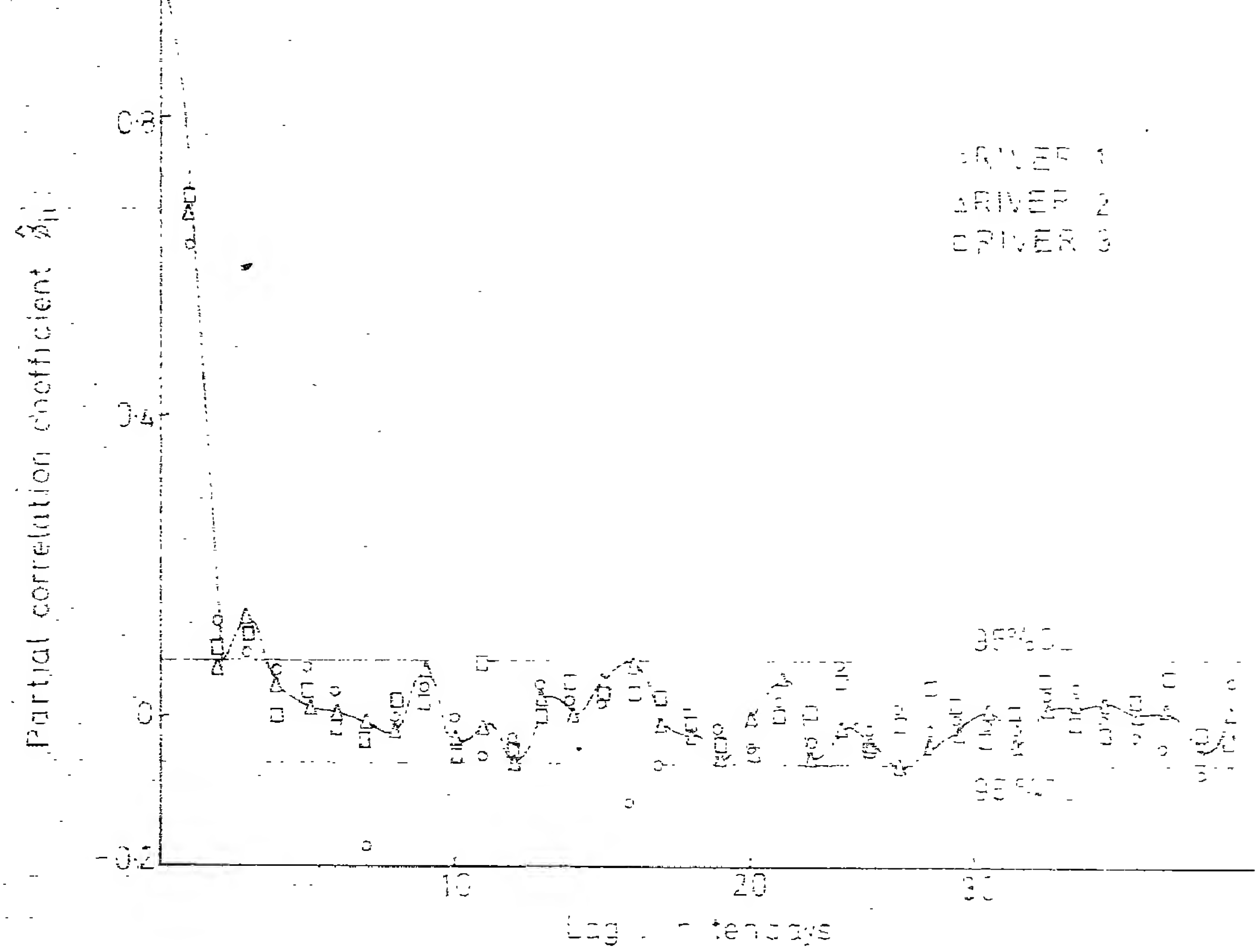


FIG. 4.5 CORRELOGRAM OF NORMALISED STANDARDISED TENDAILY SERIES



FIG·4·6 PARTIAL CORRELOGRAM OF NORMALISED STANDARDISED TENDAILY SERIES

adopted. Different values are assumed for the parameters and the sums of squares of residuals are obtained after fitting the model to the data. The sum of squared errors is plotted as a function of parameters and the region of minimum sum of squared errors is identified. A value in this region is used as a starting point. The initial estimates are inserted into an algorithm that is iterative in nature and the final estimates are calculated. The method of nonlinear least squares regression (Box and Jenkins, 1970) is used for this purpose. Program listings developed by McKerchar et al., (1972) were used to find out the final estimates of the parameters as well as the standard errors of estimates of these parameters.

To test the significance of the parameter estimates the following procedure is adopted: The parameter estimates are assumed to be jointly normally distributed with mean value equal to the true parameter values and variance-covariance matrix given by (Nelson, 1973)

$$V(B) = 2\sigma^2(u) \begin{bmatrix} \frac{\partial^2 S(B)}{\partial B^2(1)} & \cdots & \cdots & \frac{\partial^2 S(B)}{\partial B(1)\,\partial B(p+q+1)} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 S(B)}{\partial B(p+q+1)\,B(1)} & & \cdots & \frac{\partial^2 S(B)}{\partial B^2(p+q+1)} \end{bmatrix} \quad (4.9)$$

where B is the vector of the p + q + 1 parameters of ϕ and θ; S(B) is the sum of squared errors; $\sigma^2(u)$ is estimated by the sample variance and second derivatives of S(B) are numerically evaluated at the points of the parameters estimates. The square roots of the diagonal elements of the variance-covariance matrix give the standard errors SE(B(i)). Under the hypothesis that the true parameter values B(i) are zero, B(i)/SE(B(i)) is approximately normally distributed so that a 95% confidence limit for B(i) is given by -1.96 SE(B(i)) $\leq$ B(i) $\leq$ 1.96 SE(B(i)).

Parameter estimates for monthly and tendaily flows. AR models were identified in Subsec 4.1.1. The initial estimates of the parameters were made using Yule-Walker equations. From the values of the serial correlation coefficients at various lags, the above equations were solved for the initial estimates. When these values were inserted in the iterative algorithm, the final estimates of the parameters converged rapidly. The convergence was quicker for the monthly series than for the tendaily series. In fact as the process is purely AR, initial estimates by Yule-Walker equations were not necessary at all for convergence and the initial estimate of zero yielded a fast convergence towards the final values. The estimates of the parameters together with their standard errors are given in Table 4.1 for the monthly flows for a first order AR model and in Table 4.2 for tendaily flows for AR models of orders 1, 2 and 3.

TABLE 4.1 PARAMETER ESTIMATES FOR MONTHLY SERIES (STATIONARY UNIVARIATE AR MODEL)

RIVER	ORDER OF MODEL	COEFFI-CIENT	INITIAL ESTIMATE	FINAL ESTIMATE	STANDARD ERROR OF ESTIMATE	RESIDUAL VARIANCE
1	1	ϕ_1	0.52	0.5176	0.0495	0.7330
2	1	ϕ_1	0.57	0.5714	0.0475	0.6747
3	1	ϕ_1	0.56	0.5593	0.0493	0.6916

4.1.3 Validation of the fitted model

The adequacy of the model is checked through diagnostic checking (Box and Jenkins, 1970) by an examination of the characteristics of the residual series. Univariate modelling is based on the assumption that the residual series is serially independent. Furthermore for an AR process with a normal marginal distribution, the residuals should also be normally distributed. Hence the fitted model is considered adequate if each residual series constitutes an independent normally distributed series. It should be noted that the independence is necessary while normality is desirable. On the basis of the parameter estimates already made, the residual series is calculated and tested for normality and independence.

Normality of Residuals: Chisquare test was used to test the goodness of fit of normal distribution to the residual series. The test was done separately for each season.

TABLE 4.2 PARAMETER ESTIMATE FOR TENDAILY SERIES (STATIONARY UNIVARIATE AR MODEL)

RIVER	ORDER OF MODEL	COEFFICIENTS	INITIAL ESTIMATE	FINAL ESTIMATE	STANDARD ERROR OF ESTIMATE	RESIDUAL VARIANCE
1	1	ϕ_1	0.62	0.6219	0.0261	0.6174
	2	ϕ_1	0.53	0.5352	0.0331	0.6033
		ϕ_2	0.14	0.1397	0.0330	
	3	ϕ_1	0.52	0.5240	0.0333	0.6002
		ϕ_2	0.09	0.0972	0.0375	
		ϕ_3	0.09	0.0796	0.0333	
2	1	ϕ_1	0.67	0.6739	0.0246	0.5467
	2	ϕ_1	0.63	0.6340	0.0332	0.5455
		ϕ_2	0.06	0.0592	0.0332	
	3	ϕ_1	0.62	0.6265	0.0333	0.5371
		ϕ_2	-0.03	-0.0223	0.0391	
		ϕ_3	0.14	0.1285	0.0331	
3	1	ϕ_1	0.67	0.6723	0.0247	0.5481
	2	ϕ_1	0.61	0.6145	0.0333	0.5448
		ϕ_2	0.09	0.0860	0.0333	
	3	ϕ_1	0.60	0.6054	0.0332	0.5392
		ϕ_2	0.03	0.0205	0.0388	
		ϕ_3	0.10	0.1066	0.0332	

(i) Monthly series: The results of the chisquare test on the residuals are given in Table 4.3. The results of the chisquare test for the monthly series are given in Table 3.3.

TABLE 4.3 χ^2 ESTIMATES OF UNIVARIATE MONTHLY RESIDUALS (STATIONARY MODEL)

MONTH	1	2	3	4	5	6	7	8	9	10	11	12
					RIVER 1							
$\chi^2_{ESTIMATE}$	1.3	2.1	2.6	2.6	19.6	4.4	10.5	1.3	2.3	6.1	2.6	2.5
DOF	3	3	3	3	3	3	3	3	3	3	3	3
$\chi^{2+}_{THEORETICAL}$	7.8	7.8	7.8	7.8	7.8	7.8	7.8	7.8	7.8	7.8	7.8	7.8
					RIVER 2							
$\chi^2_{ESTIMATE}$	1.2	0.8	1.4	3.1	8.2	4.0	1.6	1.5	0.3	3.1	1.0	4.6
DOF	3	3	3	3	3	3	3	3	3	3	3	3
$\chi^{2+}_{THEORETICAL}$	7.8	7.8	7.8	7.8	7.8	7.8	7.8	7.8	7.8	7.8	7.8	7.8
					RIVER 3							
$\chi^2_{ESTIMATE}$	5.2	3.4	2.2	1.3	14.7	5.7	1.8	7.0	3.7	7.8	4.2	1.9
DOF	4	4	4	4	3	4	3	4	4	3	4	4
$\chi^{2+}_{THEORETICAL}$	9.5	9.5	9.5	9.5	7.8	9.5	7.8	9.5	9.5	7.8	9.5	9.5

+ at 95% confidence level

A comparison of the two indicates that the fit of the normal distribution to the residuals for seasons 5 and 7 for river 1 and for season 5 for rivers 2 and 3 are not good. For river 1 during season 5 the fit was not good for the original data as well as for the residuals. The fit was good in all the other cases.

(ii) Tendaily series: The results of chisquare test for the goodness of fit of a normal distribution for the residuals of AR models of orders 1,2 and 3 are summarised in Table 4.4.

TABLE 4.4 RESULTS OF χ^2 TEST ON UNIVARIATE TENDAILY RESIDUALS (STATIONARY MODEL)

RIVER	NUMBER OF SIGNIFICANT χ^2 VALUES		
	I ORDER	II ORDER	III ORDER
1	5	6	8
2	3	5	6
3	5	6	6

A comparison of Table 4.4 with Table 3.4(b) indicates that the goodness of fit is not satisfied for a much larger number of cases in the case of the residuals than in the case of the logtransformed series. Furthermore there is no consistency with reference to the seasons in that for some seasons the fit was good for the logtransformed series

and not for the residuals and vice versa for some others. The detailed results of the chisquare test are not given here. They indicated that the goodness of fit was always accepted for the monsoon season for the original data as well as the residuals and was rejected generally for some nonmonsoon seasons. They further indicated that the seasons in which the test for goodness of fit was rejected for the residuals clustered around the seasons for which the goodness of fit was rejected for the logtransformed data.

Independence of residuals: Univariate modelling of the data can be regarded as an attempt to find a transformation that removes the internal persistence from the data, viz., the residuals should be serially uncorrelated. Hence the correlogram of the residuals is tested to check whether the residuals are serially independent. The tests described in Subsec. 4.1.1 can be used for this purpose.

The test for independence of residuals can in addition be performed in the frequency domain. If the residual series is pure random, then the power spectra of the residuals correspond to white noise spectra. For the latter, the spectral density over the frequency range is constant and is the average of the computed values of the spectra. The spectral density function is estimated by calculating the autocorrelation function of the series and then

Fourier-transforming it. The sample spectral estimates are distributed about the population spectrum according to a χ^2/ν distribution, where ν, the number of degrees of freedom is estimated according to

$$\nu = \frac{2N}{m} - \frac{2}{3} \tag{4.10}$$

where m stands for the number of lags used for the estimation of correlation coefficients. From statistical tables, $\chi^2(\frac{\alpha}{2}, \nu)$ and $\chi^2(1 - \frac{\alpha}{2}, \nu)$ are read for a given level of significance α. From these values, the values of χ^2/ν are readily determined. If G(f) denotes the spectral density for pure random process and $G_K(f)$ that for the given sample, then $G_K(f)/G(f)$ has a χ^2/ν distribution and the series is pure random at confidence level $(1-\alpha)$ if

$$\frac{\chi^2(\frac{\alpha}{2}, \nu)}{\nu} \leqslant \frac{G_K(f)}{G(f)} \leqslant \frac{\chi^2\ (\frac{\alpha}{2}, \nu)}{\nu} \tag{4.11}$$

There are a number of other procedures including nonparametric methods for testing independence. They may be obtained from reference (Wallis et al., 1971).

Monthly residuals: The residuals for the monthly series for the first order AR model were tested for serial independence. Treating the residuals as a stationary series, the correlogram was estimated upto a lag of 60 (Fig. 4.7).

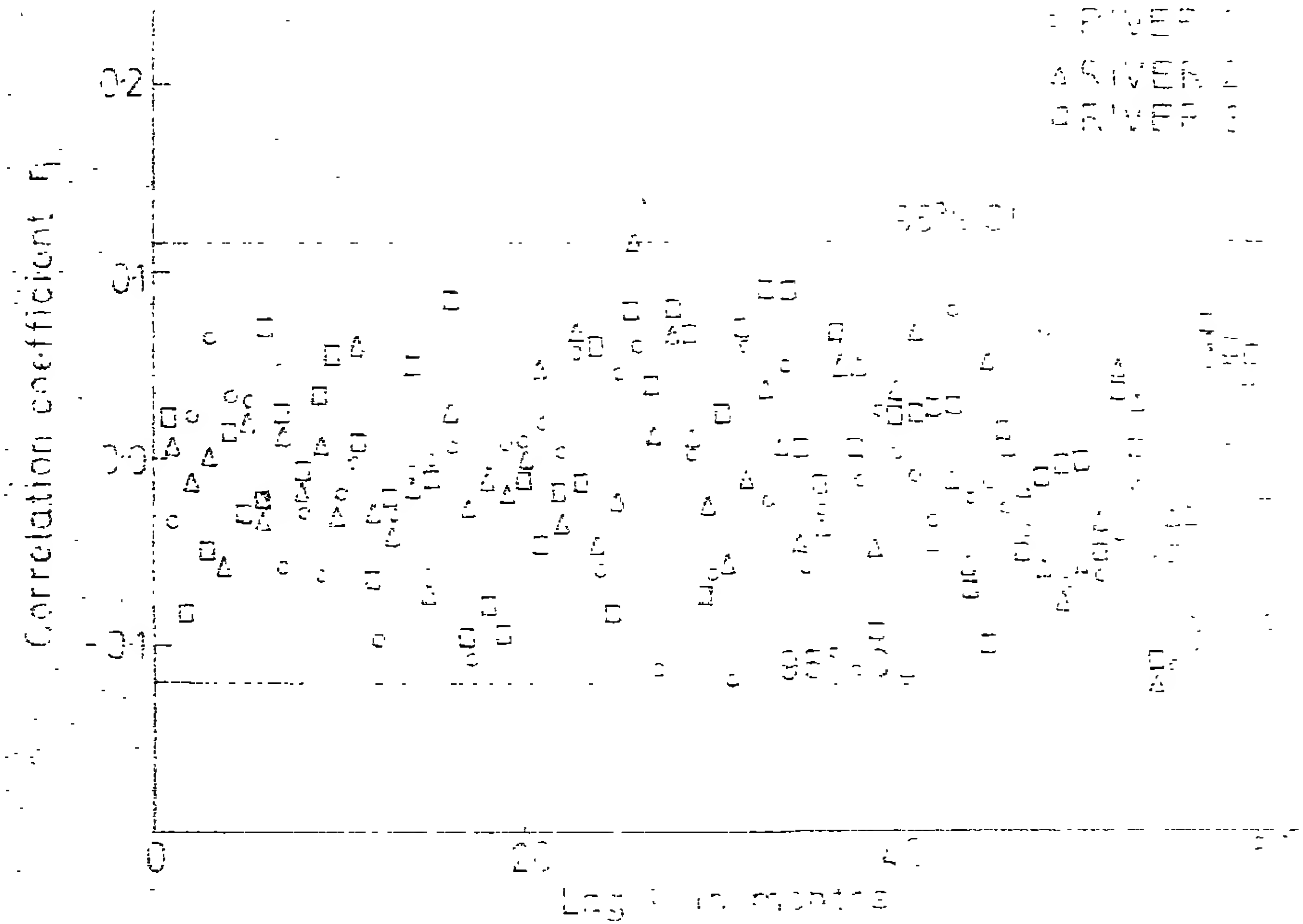


FIG.4.7 CORRELOGRAM OF MONTHLY RESIDUALS
STATIONARY FIRST ORDER AR MODEL

Assuming normality, the hypothesis that the correlogram ordinates are not significantly different from zero was tested at 95% level. All values were lying within the 95% confidence limits. In addition the Q statistic was also estimated (Table 4.5) and found to be not significant thus confirming independence. When the spectra of residuals were tested for randomness (Fig. 4.8), they were also found to be within the 95% confidence limits for all the three rivers. Hence the monthly residuals for all the three rivers can be considered to be random series.

Tendaily residuals: The correlogram of the residuals of the first, second and third order models (Figs. 4.9, 4.10, and 4.11) showed that 6 to 10 percent of the values exceeded the 95 confidence limits, and that the third order model is better than the first and second order models. When the Q statistic was computed (Table 4.5) the relative superiority of each of the models was noticed. As the order of the model increased, the statistic decreased in magnitude. However, for the first and second order models, these values were higher than the allowable values. For the third order model, the above hypothesis was satisfactory for rivers 2 and 3 and for river 1, the statistic was marginally higher than the permissible value. A larger lag model may eliminate this problem, but was not considered necessary for thisstudy.

TABLE 4.5 Q-STATISTICS OF UNIVARIATE RESIDUALS (STATIONARY MODELS)

ORDER OF MODEL		MONTHLY SERIES RIVER 1	RIVER 2	RIVER 3
1	Q	35.5	15.8	33.2
	DOF	29	29	29
	$\chi^2_{95\%}$	42.6	42.6	42.6

ORDER OF MODEL		TENDAILY SERIES RIVER 1	RIVER 2	RIVER 3
1	Q	58.4^+	48.2^+	59.3^+
	DOF	29	29	29
	$\chi^2_{95\%}$	42.6	42.6	42.6
2	Q	52.5^+	44.3^+	54.0^+
	DOF	28	28	28
	$\chi^2_{95\%}$	41.3	41.3	41.3
3	Q	40.8^+	27.8	38.3
	DOF	27	27	27
	$\chi^2_{95\%}$	40.1	40.1	40.1

\+ denotes significance at the 95% level.

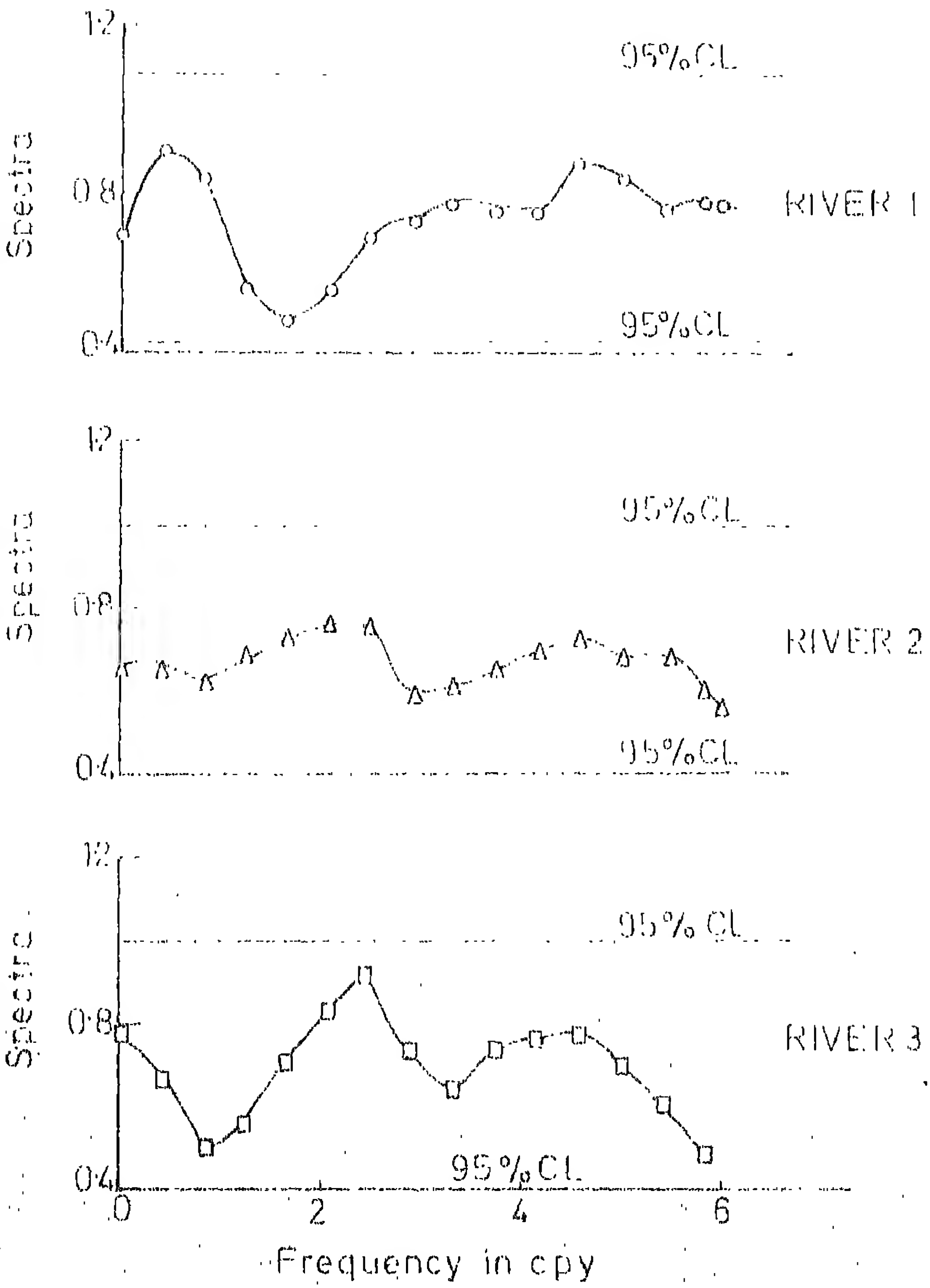
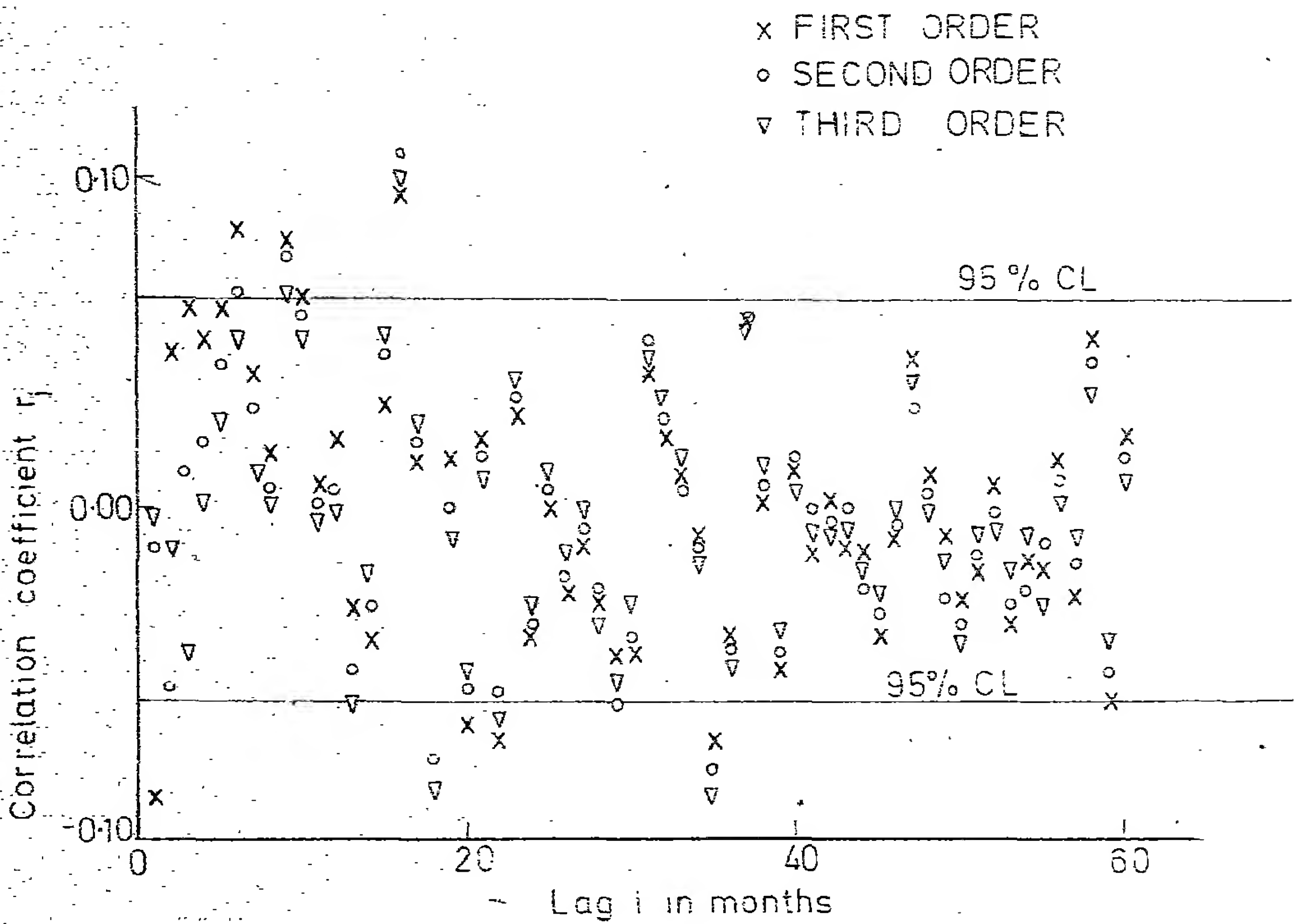


FIG. 4·8 SPECTRA OF MONTHLY RESIDUALS OF STATIONARY FIRST ORDER AR MODEL



FIG·4·9 CORRELOGRAM OF TENDAILY RESIDUALS OF STATIONARY AR MODEL, RIVER 1

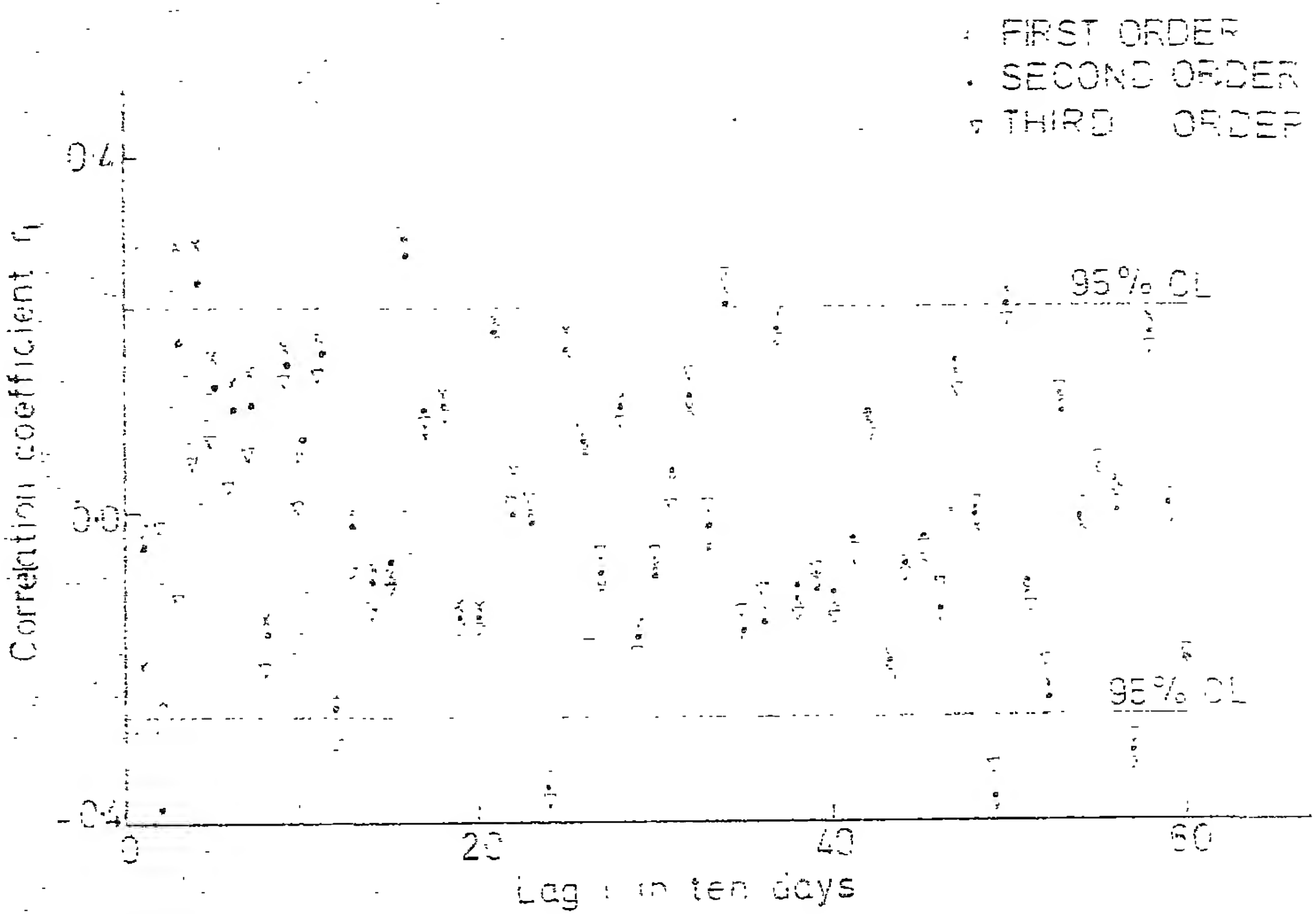


FIG. 4.10 CORRELOGRAM OF TENDAILY RESIDUALS OF STATIONARY AR MODELS: RIVER 2

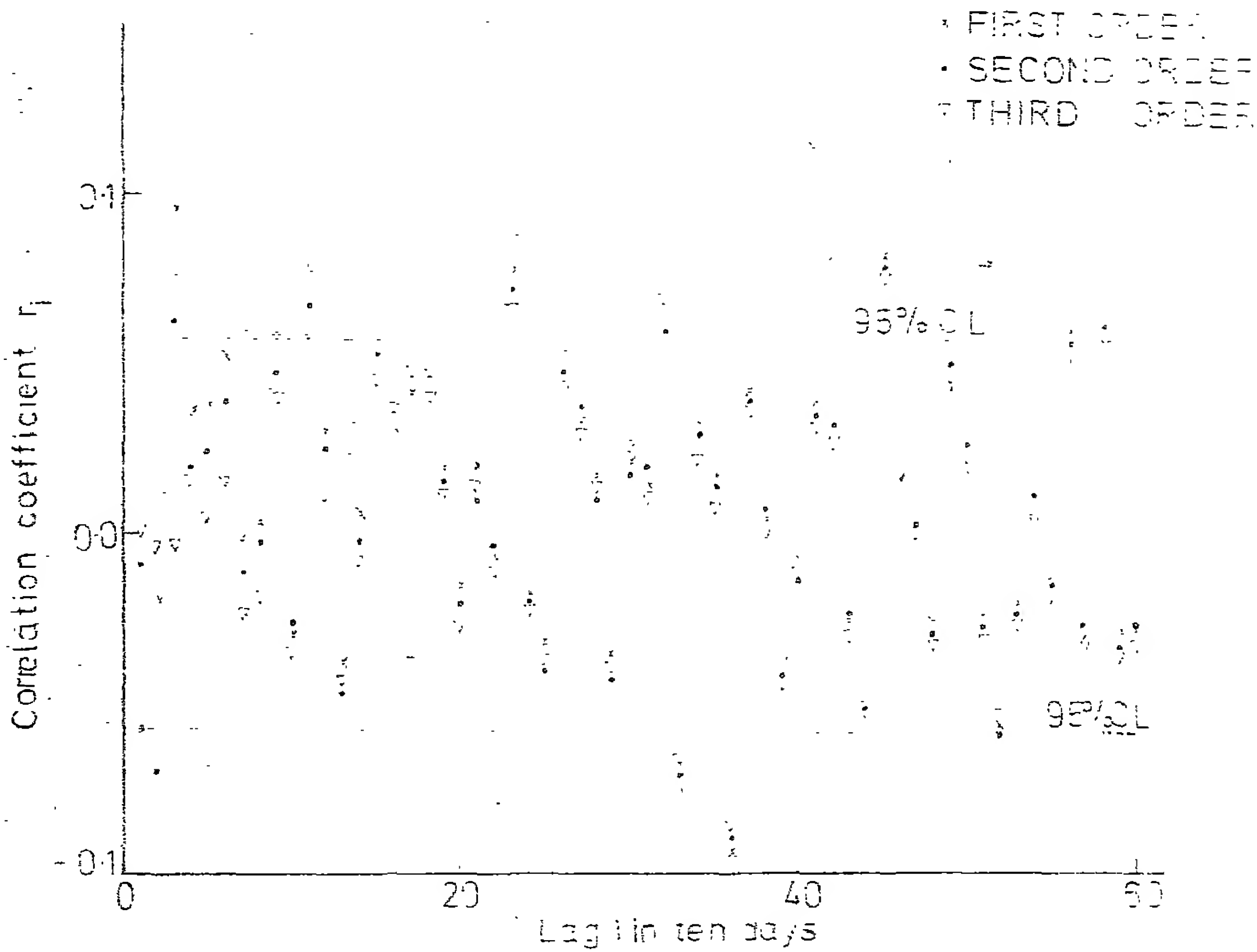
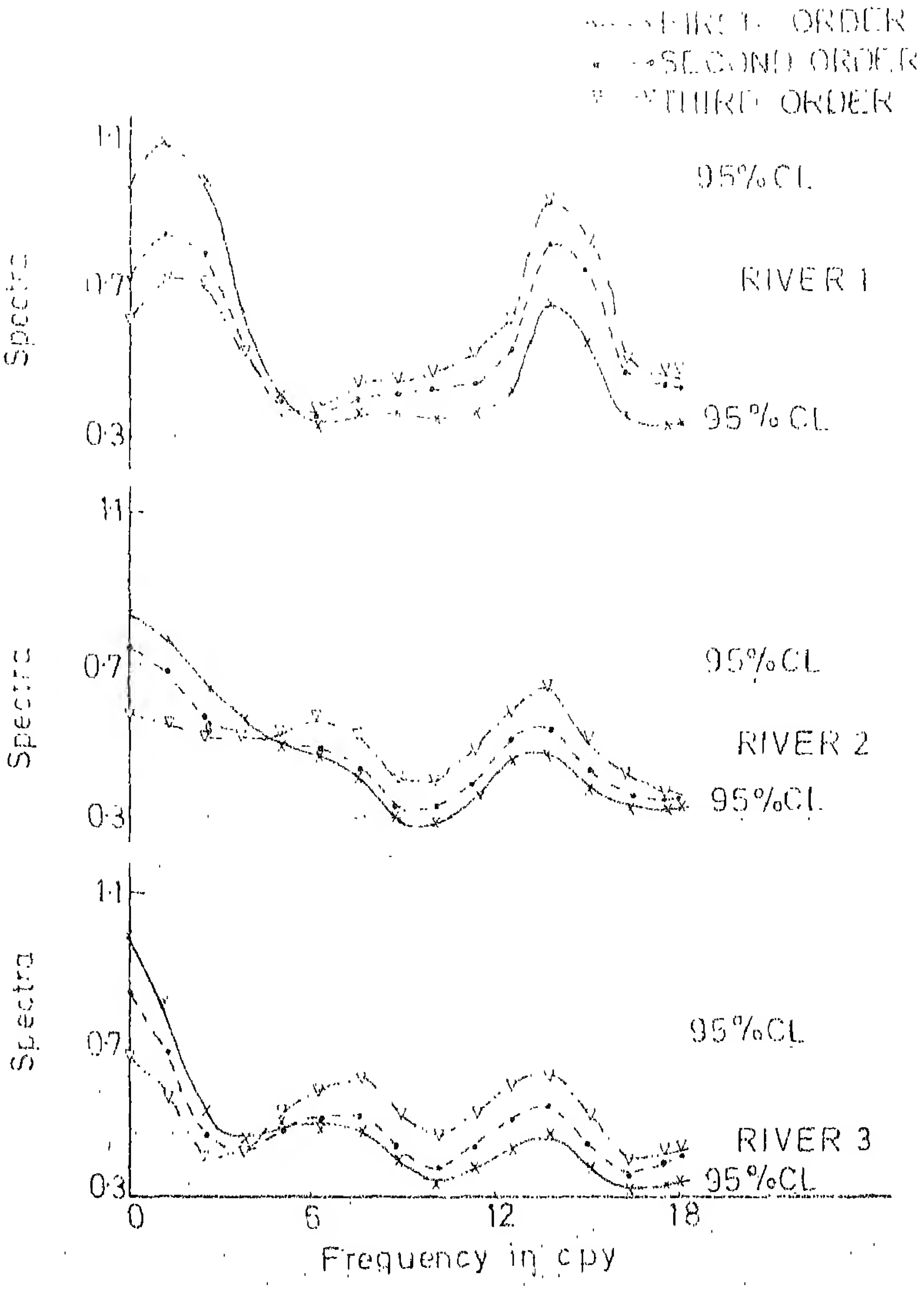


FIG. 4.11 CORRELOGRAM OF TENDAILY RESIDUALS OF STATIONARY AR MODELS, RIVER 3

The spectra of the tendaily residuals are plotted in Fig. 4.12 along with the confidence band for a white noise spectra and they indicate that the first and second order models are not adequate to represent persistence in the series and that the third order model is necessary. Hence it was decided to adopt the third order model for the tendaily series.

Seasonwise independence of residuals: The correlogram analysis of the residuals of the stationary AR model enables the testing of the independence of the residuals as a whole, i.e., treated as a single stationary series. Whether from season to season the residuals satisfy the condition of serial independence is not revealed by the correlogram analysis. The latter can be calculated after grouping the residuals according to seasons and estimating the correlation between the residuals belonging to a particular season and those belonging to the immediately earlier season. This is referred to as the 'seasonal serial correlation coefficient'. For instance, considering the monthly residual series, the correlation coefficient between the residuals of January and the previous month, December, may be estimated and tested for the significance of the residual correlation. The hypothesis that the population correlation coefficient between two seasonal residuals is zero is set up. The t statistic is



FIG·4·12 SPECTRA OF TENDAILY RESIDUALS OF STATIONARY AR MODELS

calculated using

$$t = \frac{r_1^j \sqrt{N-2}}{\sqrt{(1-(r_1^j)^2)}} \tag{4.12}$$

where N is the number of pairs of the residuals in the sample and r_1^j is the sample correlation coefficient between the j-th and (j-1)-th seasons. If the calculated t statistic is less than the tabular value at the given level of significance for (N-2) degrees of freedom, the hypothesis is accepted.

Seasonal variation does not arise for the annual series. Seasonwise serial correlation coefficients for the residuals of the monthly series are shown in Fig. 4.13 and for the tendaily series in Figs. 4.14, 4.15 and 4.16. It can be seen that the coefficient is significant for the monthly series only on one occasion for rivers 2 and 3. For the tendaily series, the number of such instances are greater. For the first order model, the residuals possessed significant correlations between adjacent seasons for 7, 3 and 7 seasons for rivers 1, 2 and 3 respectively. For the second order model, these figures were 9, 4 and 8 showing an increase for all the rivers. For the third order model, rivers 1 and 3 showed an improvement inthe independence of residuals with 6 seasonal values being significant. It was thus seen that increasing theorder of autoregressive model did not always result in an improvementof the season-to-season independence

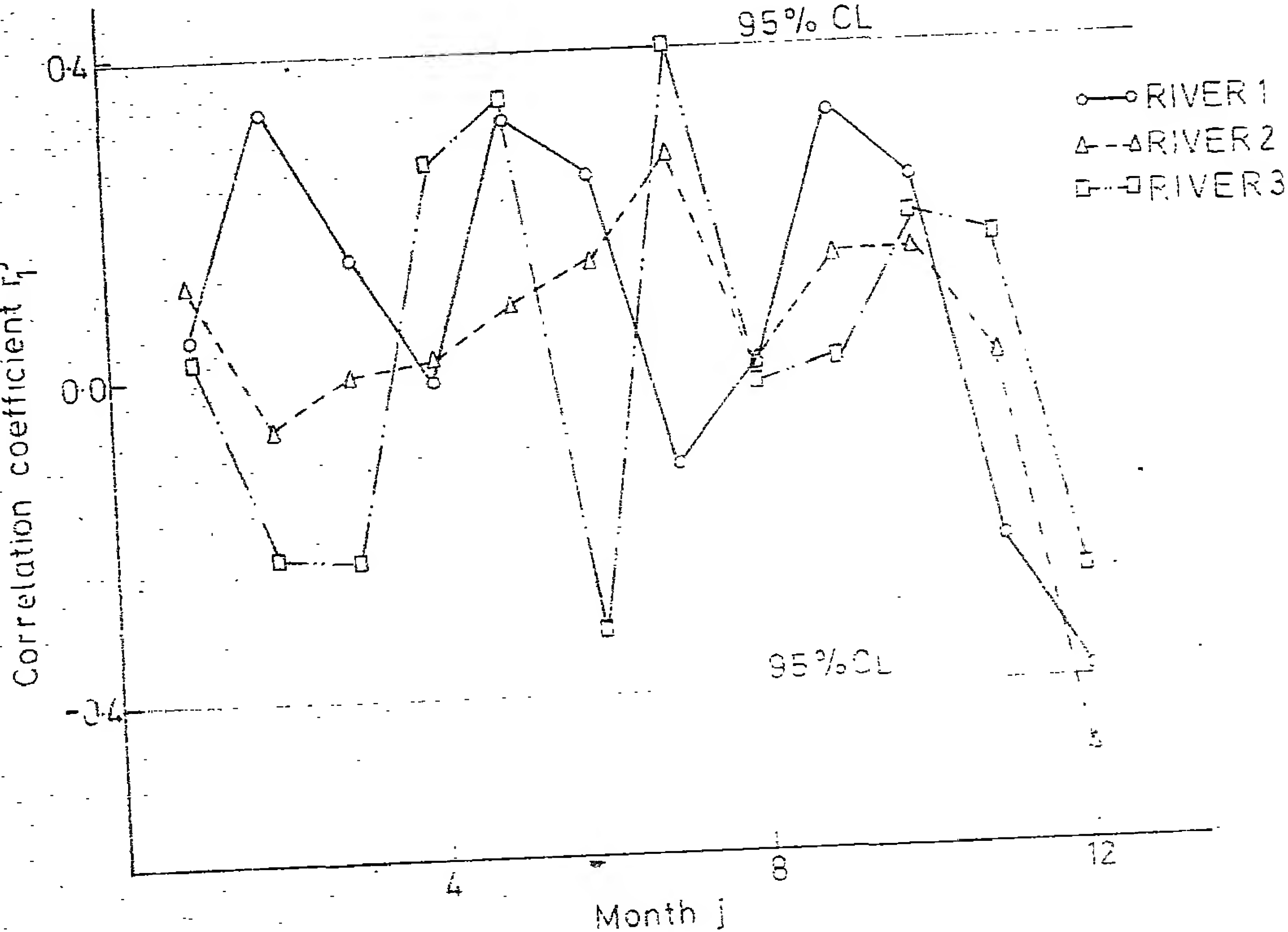
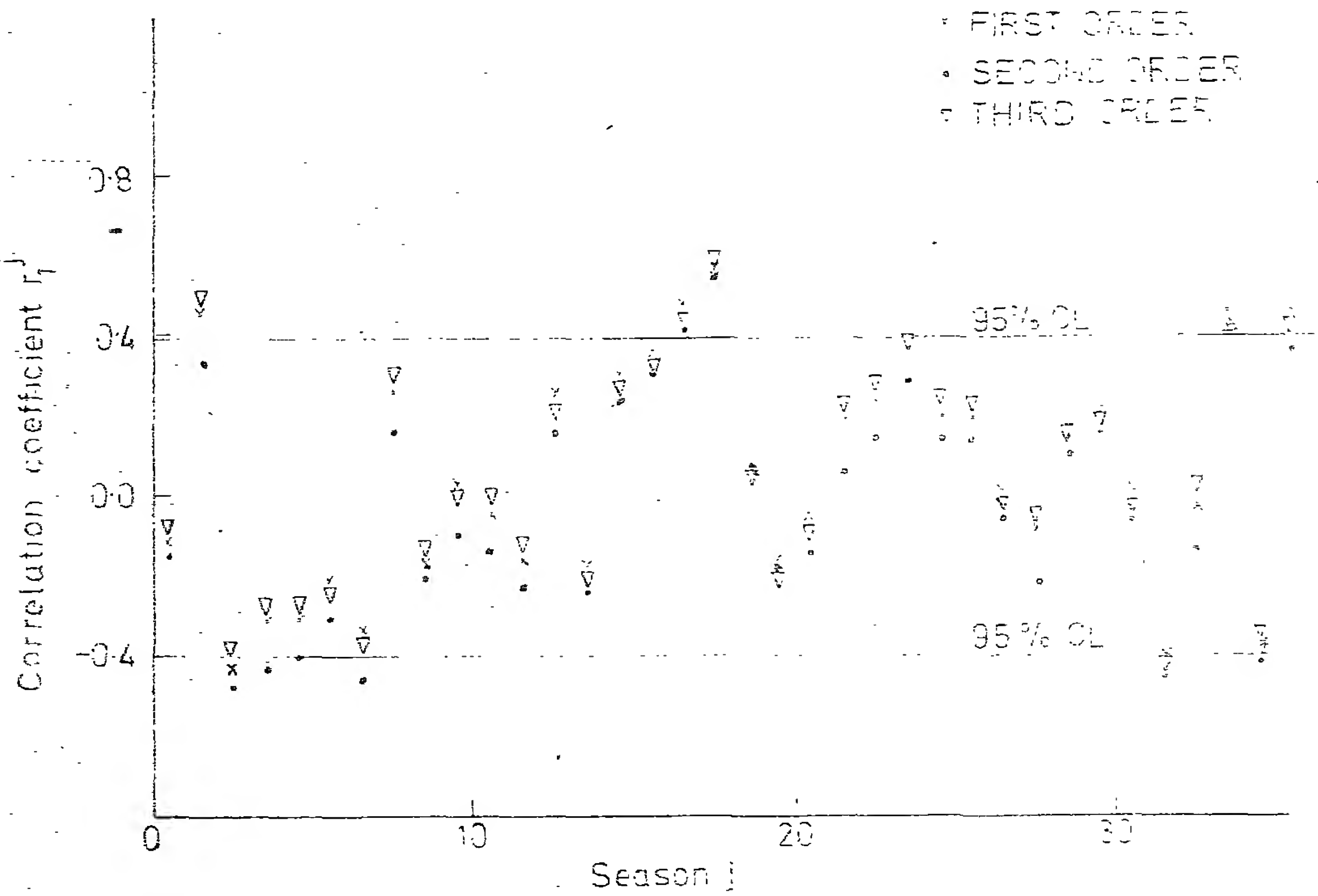
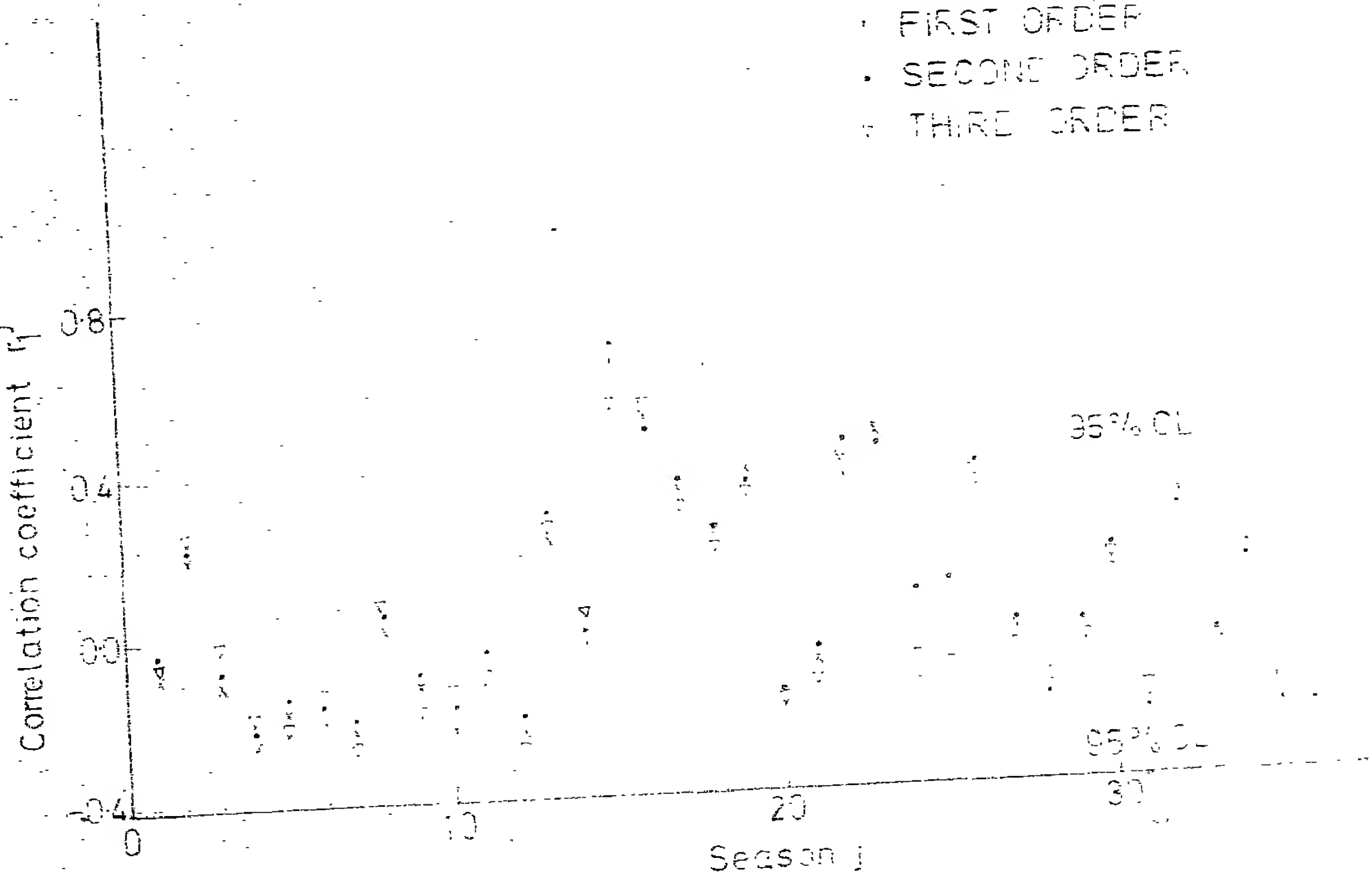


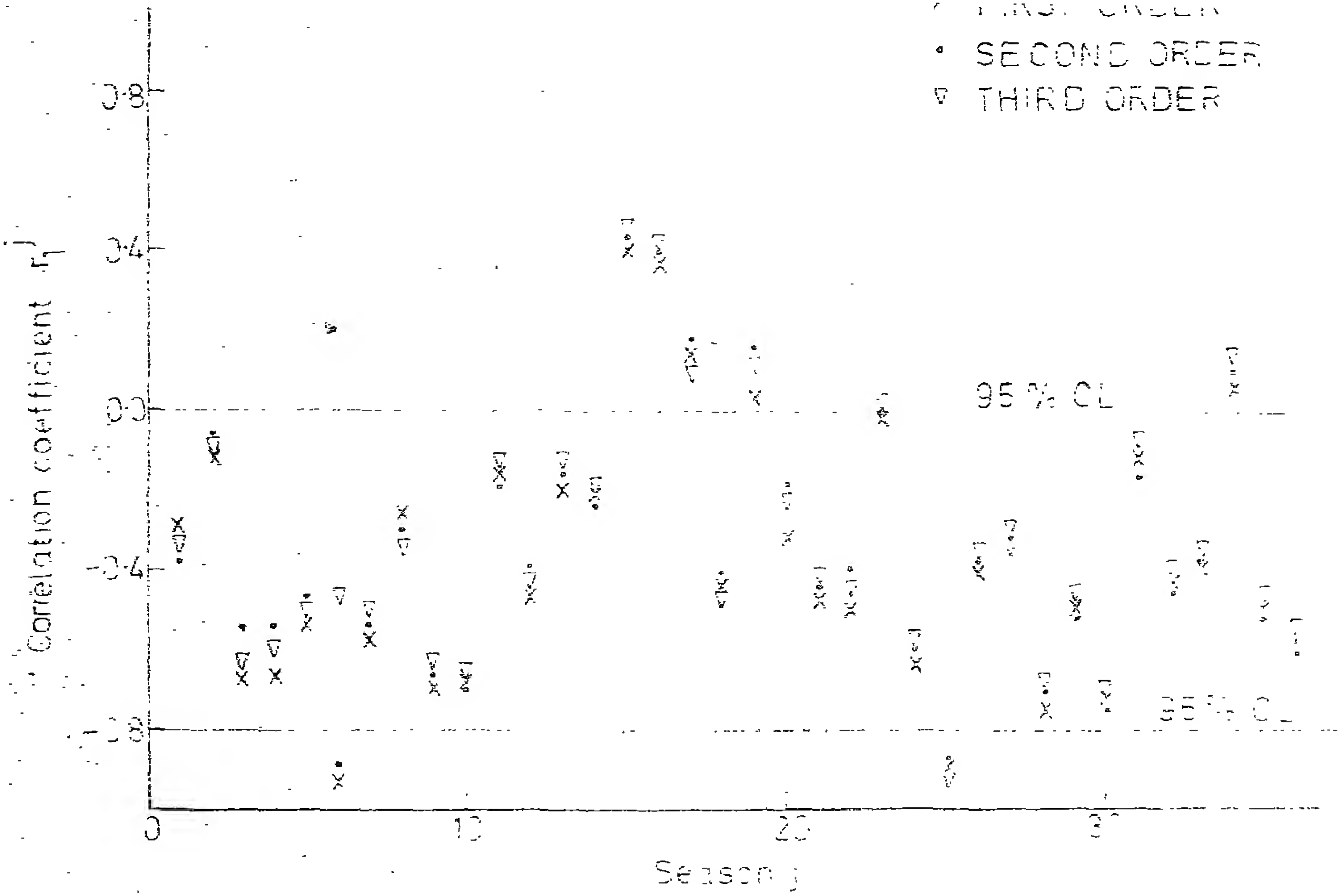
FIG. 4.13 SEASONWISE CORRELATION OF MONTHLY RESIDUALS OF STATIONARY FIRST ORDER AR MODEL



FIG·4·14 SEASONWISE CORRELATION OF TENDAILY RESIDUALS OF STATIONARY AR MODELS, RIVER 1



FIG·4·15 SEASONWISE CORRELATION OF TENDAILY RESIDUALS OF STATIONARY AR MODELS, RIVER 2



FIG·4·16 SEASONWISE CORRELATION OF TENDAILY RESIDUALS OF STATIONARY AR MODELS, RIVER 3

of the residuals. It could also be observed that during the monsoon months, the residuals did not generally have any significant seasonal correlation whereas during the autumn months, the correlations were very large and significant. The presence of significant correlation from season to season in the residuals showed that when examined on a season by season basis the model used for the univariate analysis was not able to fully account for the internal persistence during all seasons of the year, although when examined on an annual basis the model resulted in a residual series free from internal correlation. This may be due to the averaging the effects of the several seasons taken as a whole.

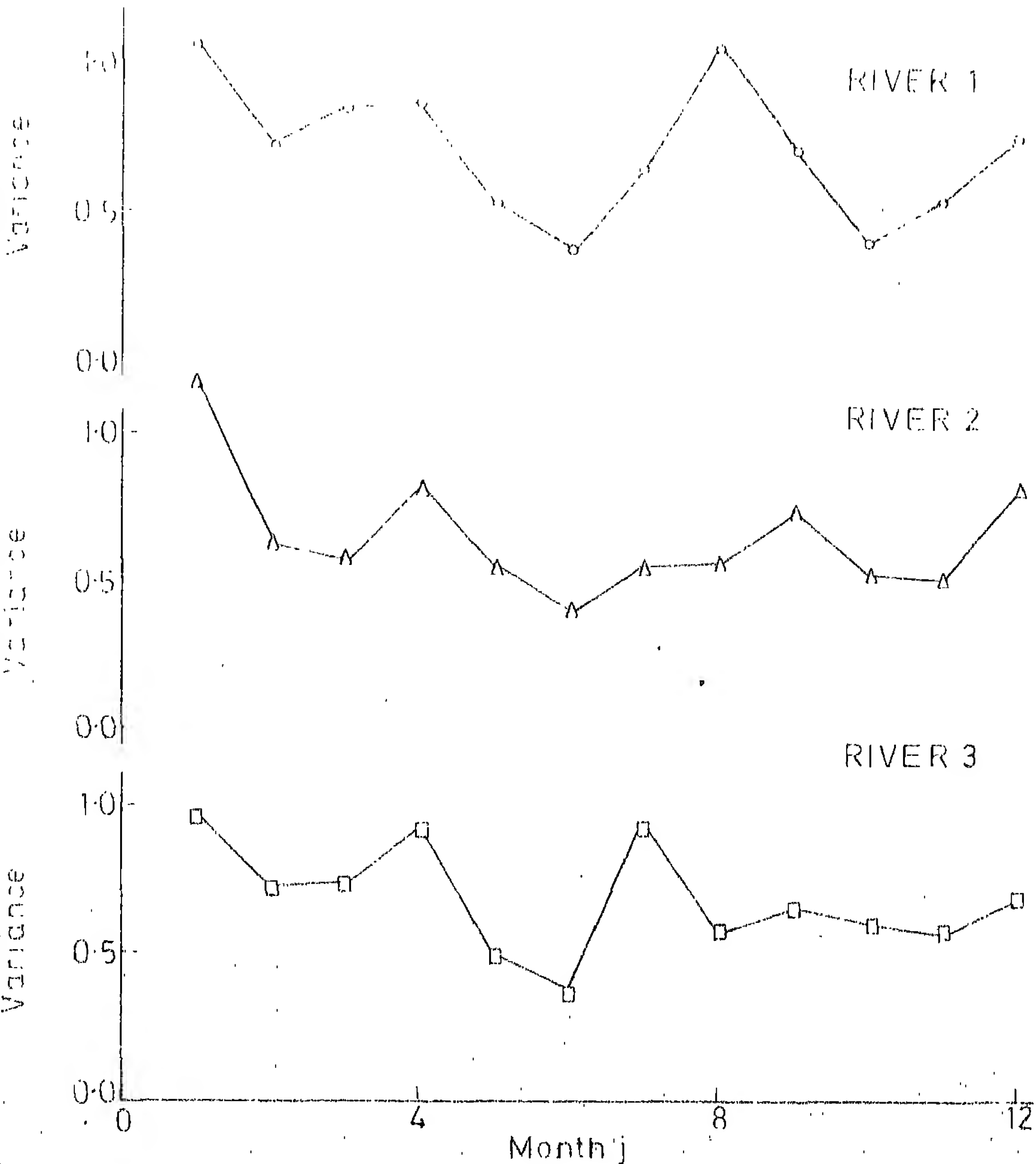
4.1.4 Variance of the residuals

For the univariate model, the total variance of the process may be divided into a variance that is accounted for by the persistence and a residual variance. The latter may be attributed to the component that is serially uncorrelated. After the model is fitted to the data the variance of the residual series can be estimated theoretically or empirically from the residuals.

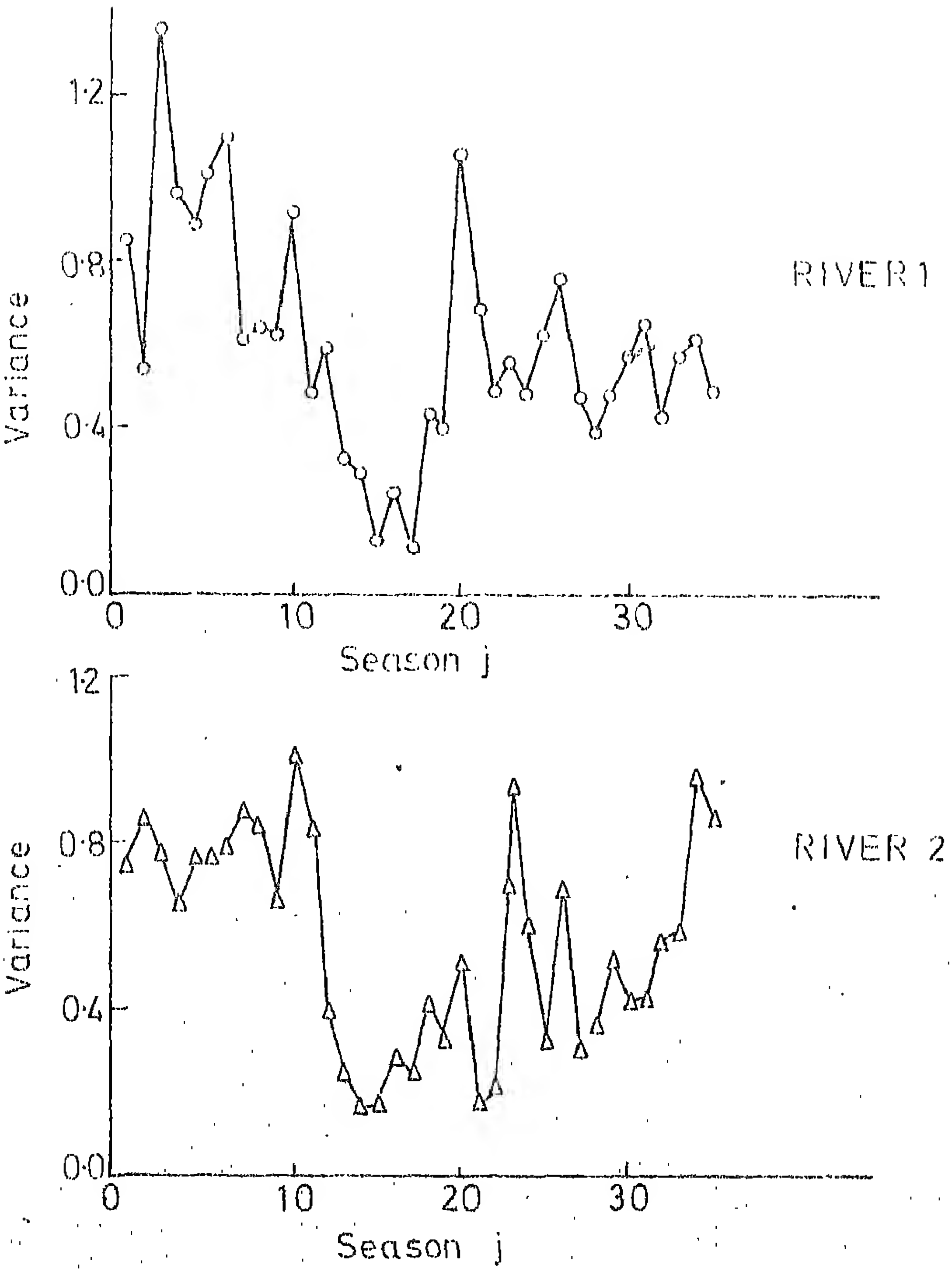
The data series have already been transformed and standardised and hence have unit variance. The variances of residuals for the monthly series for the three rivers were

respectively 0.7330, 0.6747 and 0.6916. For the tendaily series, the variances of residuals were calculated for the first, second and third order models. They are shown in Table 4.2. As the order of the model increases, there is a slight reduction in the variance. For river 1, for instance, when the order of the model increases from 1 to 2, the variance decreases from 0.6147 to 0.6003, viz., about 1.04 % of the original variance, and when the order increases from 2 to 3, there is a further reduction of 0.31% of the total variance. For rivers 2 and 3, the overall reduction in variance, is 0.96 % and 0.89 % only.to the third order model from the first order model.

In order to see how much of the persistence component was accounted for during each season, estimation of variance was made separately for each season. Residuals belonging to each season were grouped together and their variances were estimated. Results for the monthly series are shown in Fig. 4.17 and for the tendaily series (third order model) in Fig. 4.18. From season to season, the variances varied in magnitude. These values were seen to be generally higher during the monsoon season than during nonmonsoon season with a low in the autumn season. In other words, the persistence component appears to vary seasonally and the variance of the residuals estimated earlier averages out these variations



FIG·4·17 VARIANCE OF MONTHLY RESIDUALS OF STATIONARY FIRST ORDER AR MODEL



FIG·4·18 VARIANCE OF TENDAILY RESIDUALS OF STATIONARY THIRD ORDER AR MODEL

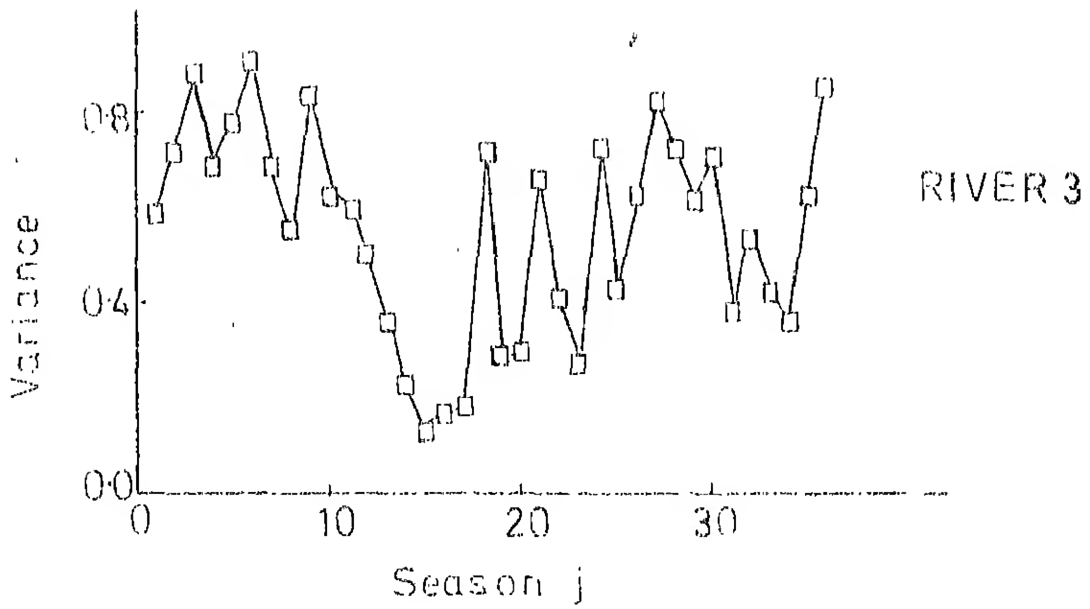


FIG: 4·18 VARIANCE OF TENDAILY RESIDUALS OF STATIONARY THIRD ORDER AR MODEL

yielding a unique value for the entire series. For some seasons, mostly during the monsoon, the residual variances were greater than the original variances. This may be due to the estimated value of the AR component being higher than the actual autoregression prevailing during monsoon. The overestimated persistence component would have resulted in a large negative value for the random component resulting in a large variance.

It is thus seen that when the residuals are treated as a single stationary series irrespective of the seasons, the variance estimates as well as the correlogram analysis seem to yield consistent results. But if the characteristics of residuals are examined on a seasonal basis, the residuals are not seasonally independent, and the variance estimates indicate nonstationarity of residuals. So a stationary univariate model is not satisfactory for the monthly and ten-daily series.

4.2 Nonstationary Models

In the modelling of monthly and tendaily series, differencing was not found to be necessary. Furthermore an AR component alone is identified and MA components are not needed. However, a stationary AR model is found to be not satisfactory. Hence it seems necessary to consider a nonstationary AR model.

4.2.1 Identification of the model

A general nonstationary model may consist of a time varying autoregressive component and a time varying moving average component and may be represented as

$$x(t) = \sum_{i=1}^{p} \phi_i^j \, x(t-i) + \sum_{i=1}^{q} \theta_i^j \, \varepsilon(t-i) + \varepsilon(t) \quad (4.13)$$

where ϕ_i^j and θ_i^j stand respectively for the autoregressive and moving average coefficients of order i for the season j, j = 1,2,... 12 for the monthly series and j = 1,2,...36 for the tendaily series. The correlogram cannot be used here for the purpose of identification because it is based on the assumption of stationarity of the correlation structure. However for the purposes of identifying the order and form of the model, assuming that they do not change from season to season, the correlogram may be used. Once the form and order of the model are identified, they are fitted into the data for the entire series. This stationary approach is followed only for the purposes of identification of the model. No parameter estimation for the model is involved.

In accordance with the results of Subsec. 4.1.1, AR models of first and third order were used respectively for monthly and tendaily series, viz., q was equal to zero and p was independent of j.

The monthly series and the tendaily series are then represented respectively by

$$x(t) = \phi_1^j\, x(t-1) + \varepsilon(t), \tag{4.14}$$

and

$$x(t) = \phi_1^j\, x(t-1) + \phi_2^j\, x(t-2) + \phi_3^j\, x(t-3) + \varepsilon(t) \tag{4.15}$$

The parameters ϕ_i^j are to be estimated for each season j for each river.

4.2.2 Estimation of parameters

Yule-Walker relationships: An extension of the Yule-Walker relationships (Eq. 4.4) is used for the solution of AR coefficients of the seasonally nonstationary time series. Let r_i^j be the i-th order seasonal correlation coefficient for j-th season, i.e., it is the correlation coefficient between the variable during the j-th season and that during the (j-i)-th season. Then r_1^j, r_2^j and r_3^j represent respectively the seasonal correlation coefficient of first second and third order for the j-th season.

For a first order AR model, i.e. for p = 1,

$$\phi_1^j = r_1^j \tag{4.16}$$

For a second order AR model, i.e., for p = 2,

$$\phi_1^j = r_1^j - r_1^j r_2^j/(1-(r_2^j)^2)$$

and

$$\phi_2^j = r_2^j - r_1^j r_2^j/(1-(r_2^j)^2) \tag{4.17}$$

For a third order AR model with $p = 3$ (Yevjevich,1972) the AR coefficients ϕ_1^j, ϕ_2^j and ϕ_3^j are respectively given by

$$\phi_1^j = [(1-(r_1^j)^2)(r_1^j-r_3^j)-(1-r_2^j)(r_1^j r_2^j-r_3^j)]/A$$

$$\phi_2^j = [(1-r_2^j)(r_2^j+(r_2^j)^2-r_1^j r_3^j)]/A \quad \text{and}$$

$$\phi_3^j = [(r_1^j-r_3^j)((r_1^j)^2-r_2^j)-(1-r_2^j)(r_1^j r_2^j-r_3^j)]/A \tag{4.18}$$

where $A = (1-r_2^j)(1-2(r_1^j)^2 + r_2^j)$.

The seasonal correlation coefficients r_i^j are estimated separately for each season j. As the data are already standardised, the seasonal correlation coefficients are numerically equal to the covariances. The i-th order seasonal correlation coefficient for the j-th season is estimated by

$$r_i^j = \frac{1}{N-1} \{x^j\}\{x^{j-i}\}^T , \tag{4.19}$$

where $\{x^j\}$ is the vector of the variables for the j-th season.

Method of least squares: The general p-th order AR model can be regarded as a multiple linear regression equation, viz., in the equation

$$x(t) = \sum_{i=1}^{p} \phi_i^j \, x(t-i) + \varepsilon(t). \tag{4.20}$$

where the coefficients ϕ_i^j may be solved by the method of least squares. The x(t) series generally has a zero mean and unit standard deviation. If standardisation is done using the seasonwise mean and standard deviation by MM or by the MLS using Chow's frequency factor, standardisation does not result in zero mean and unit standard deviation. Then the autoregressive model includes an additional term ϕ_o^j to represent the nonzero mean;

$$x(t) = \phi_o^j + \sum_{i=1}^{p} \phi_i^j \, x(t-i) + \varepsilon(t) \tag{4.21}$$

and the coefficients ϕ_i^j, $i = 0,1,2,\ldots p$; $j = 1,2,\ldots s$ are solved by the method of least squares. The solution is obtained from the following set of normal equations:

$$\begin{bmatrix} 1 & \Sigma x(t-1) & \cdots & \Sigma x(t-p) \\ \Sigma x(t-1) & \Sigma x^2(t-1) & \cdots & \Sigma x(t-1)x(t-p) \\ \cdots & & \cdots & \cdots \\ \Sigma x(t-p) & \Sigma x(t-1)x(t-p) & \cdots & \Sigma x^2(t-p) \end{bmatrix} \begin{bmatrix} \phi_o^j \\ \phi_1^j \\ \cdots \\ \phi_p^j \end{bmatrix} = \begin{bmatrix} \Sigma x(t) \\ \Sigma x(t)\, x(t-1) \\ \cdots \\ \Sigma x(t)x(t-p) \end{bmatrix} \tag{4.22}$$

For the first order model, the AR coefficients as obtained by the assumption of Yule-Walker relations and by the method of least squares yielded almost identical results. However, when the order of the model was raised, the Yule-Walker equations started giving unstable estimates, more so

for the third order model than for the second order model. This result was irrespective of whether the standardisation was done by MM, MLE or MLS. Hence in this study, the method of least squares was used for the estimation of the coefficients of the model. Estimates of the AR coefficients using standardisation by MLE and MLS are given in Table 4.6 for the monthly series. The coefficient estimates for the tendaily series for the first, second and third order AR models are not given here.

4.2.3 Validation of the Model

Normality of the residuals: The residuals were determined by substituting the estimated value of parameters in the model equation. They were grouped in accordance with the season to which they belonged. The goodness of fit of the normal distribution was verified by estimating the chi-square statistic for the residuals of each season.

(i) Monthly series: Results for goodness of fit at 95 confidence level are shown in Table 4.7. The fit was not good for 2 months out of 12 for river 1 and for rivers 2 and 3 for only one month whatever be the method of standardisation. As normality is only desirable and not essential the results were considered satisfactory for this study.

TABLE 4.6 PARAMETER ESTIMATES FOR MONTHLY SERIES (NONSTATIONARY UNIVARIATE FIRST ORDER AR MODEL)

Month	Autoregressive Coefficients								
	RIVER 1			RIVER 2			RIVER 3		
	MLE	MLS		MLE	MLS		MLE	MLS	
	ϕ_1	ϕ_o	ϕ_1	ϕ_1	ϕ_o	ϕ_1	ϕ_1	ϕ_o	ϕ_1
1	-.30	.56	.26	-.00	.17	-.01	.28	.90	.22
2	.46	1.77	.39	.58	1.47	.56	.48	1.53	.40
3	.38	.94	.17	.64	-.46	.73	.51	-.13	.53
4	.37	.27	.37	.43	.15	.43	.36	.01	.35
5	.69	-.34	.71	.68	.02	.72	.73	.31	.77
6	.85	-.02	.82	.80	-.43	.77	.88	.13	.89
7	.57	.17	.58	.67	-.07	.67	.33	-.18	.31
8	.17	-.14	.17	.65	.91	.67	.64	-.27	.70
9	.51	-.14	.50	.50	.12	.51	.59	.19	.53
10	.83	.13	.82	.69	-.46	.68	.63	-.49	.67
11	.78	-.23	.79	.71	.38	.70	.66	.12	.64
12	.67	.59	.67	.44	.43	.46	.56	.28	.61

TABLE 4.7 RESULTS OF χ^2 TEST ON UNIVARIATE RESIDUALS (NONSTATIONARY MODELS)

NUMBER OF OCCASIONS OF SIGNIFICANT χ^2 ESTIMATES

RIVER	MONTHLY SERIES	TENDAILY SERIES		
	ORDER	ORDER		
	1	1	2	3
1	2	9	9	8
2	1	2	2	2
3	1	4	4	3

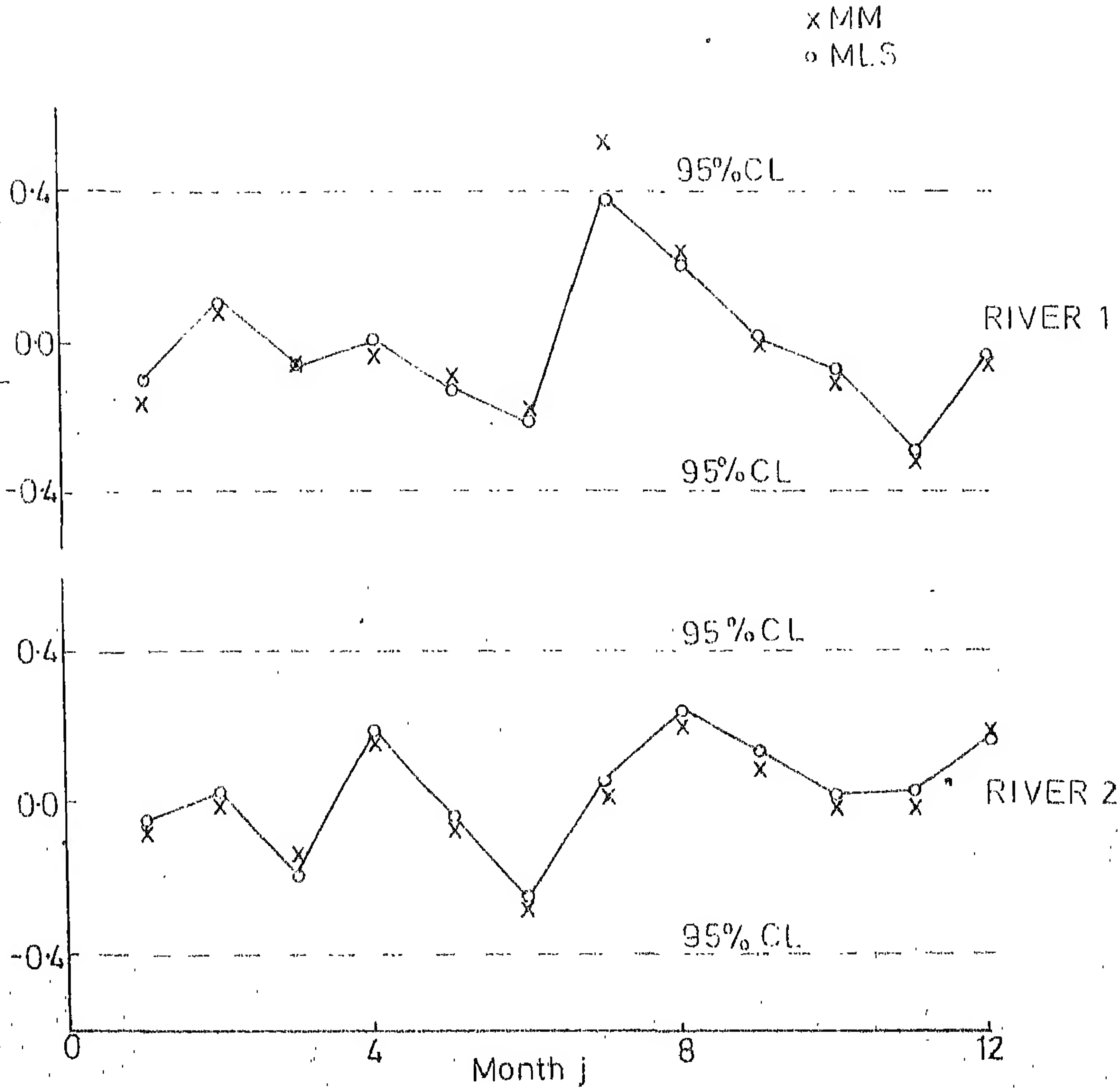
(ii) Tendaily series: The residuals of the first, second and third order models were determined for the tendaily series of each river. 36 groups of tendaily residual series corresponding to each season were identified and individually tested for normality. It was noted that the numbers of occasions when fit was not good were same for all methods of standardisation of data. As the order of the model was increased, from first to second, there was no improvement in fit. But, for the third order model, rivers 1 and 3 showed a marginal improvement. It is likely that by using a higher lag model, the goodness of fit may be increased. This would involve the estimation of more number of coefficients. For a particular length of data, then, reliability of the estimates may decrease. Furthermore, as stated earlier, normality is not essential. Hence the results were considered satisfactory for the model used in this study.

Independence of the residuals: The serial independence of the residuals was tested by correlogram and spectral analysis and the results were satisfactory for both the monthly and tendaily series. The seasonal independence of the residuals was checked for the monthly and tendaily series using the procedures of Subsec 4.1.3.

It was found that for all months and for all the methods of standardisation, except for one month each for rivers 1 and 3 using MM, the fit was good (Fig. 4.19). The seasonwise serial correlation coefficients of the residuals of the tendaily series for the first, second and third order models are shown in Figs. 4.20 to 4.26 (for MLS only). For the first order model, the residuals have significant serial correlation coefficients for 3, 2 and 2 seasons respectively for rivers 1, 2 and 3. For the higher order models, there are no significant seasonal serial correlation coefficients except when the method of least squares is used for river 2 (2 seasons). This is within 5% of the total number of seasons. Hence it can be concluded that provided a second or third order model is used for the tendaily series, the residuals are serially and seasonally uncorrelated whatever be the method of standardisation.

4.2.4 Variance of residuals

(i) Monthly series: Variances were estimated for the seasonal residuals of the first order model (Figs. 4.27 to 4.29). Of the three methods used for standardisation, MLS appears to give a lower estimate of the variance than the other two methods. The other two methods seem comparable to each other. Late autumn, early winter and summer seem to be the seasons with minimum unexplained variance and it is during



FIG·4·19 SEASONWISE CORRELATION OF MONTHLY RESIDUALS OF NONSTATIONARY FIRST ORDER AR MODEL

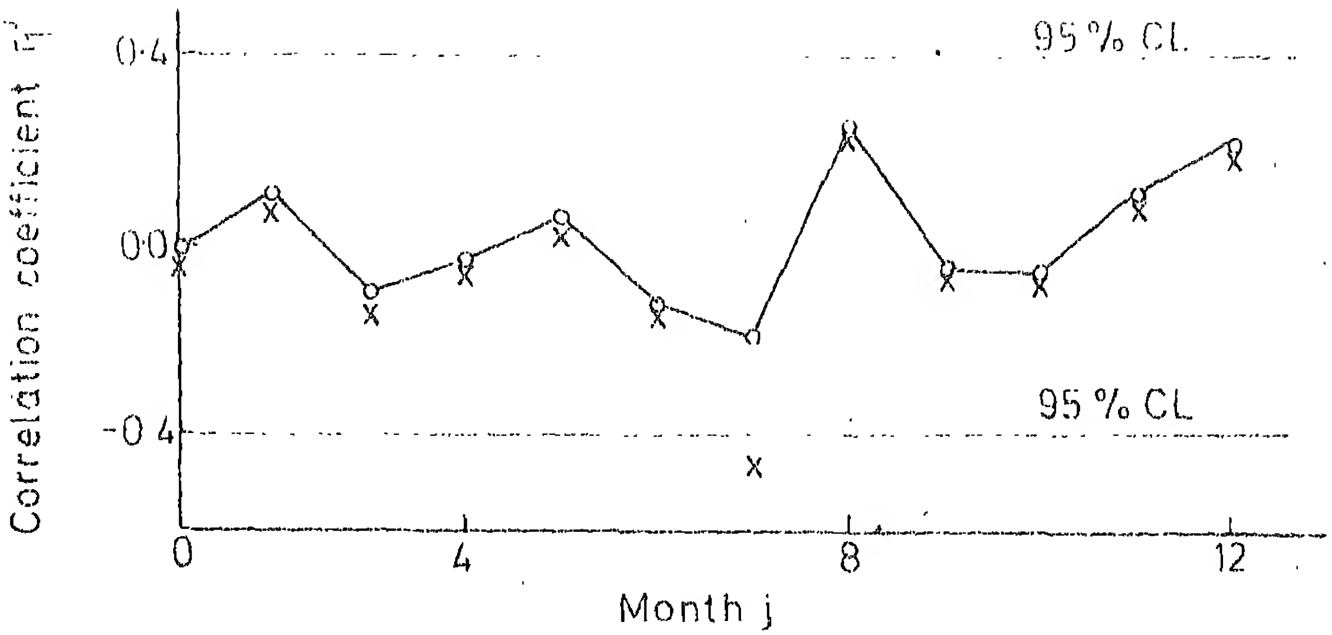


FIG. 4.19 SEASONWISE CORRELATION OF MONTHLY RESIDUALS OF NONSTATIONARY FIRST ORDER AR MODEL (CONT.)

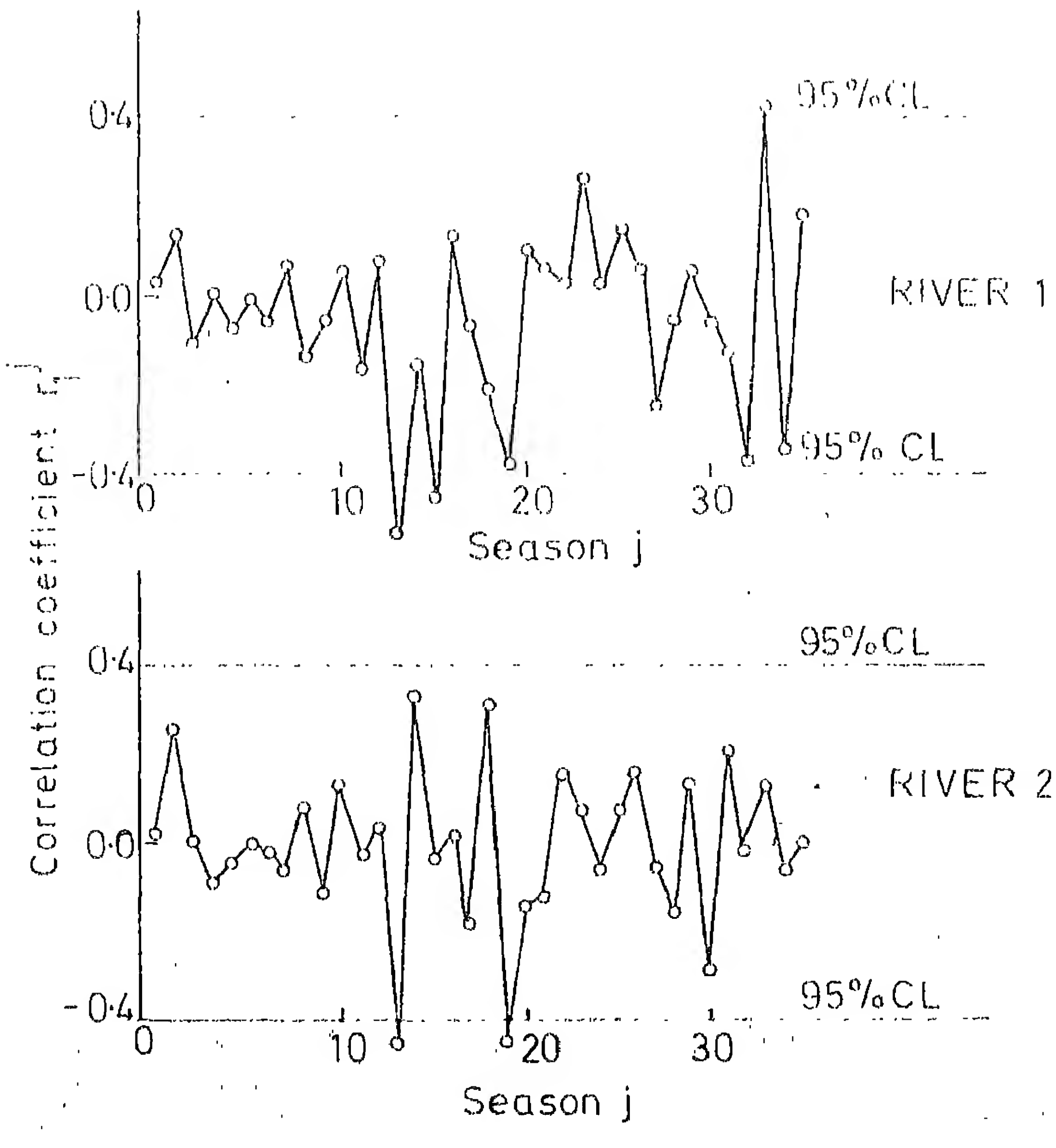


FIG. 4.20 SEASONWISE CORRELATION OF TENDAILY RESIDUALS (MLS) OF NONSTATIONARY FIRST ORDER AR MODEL

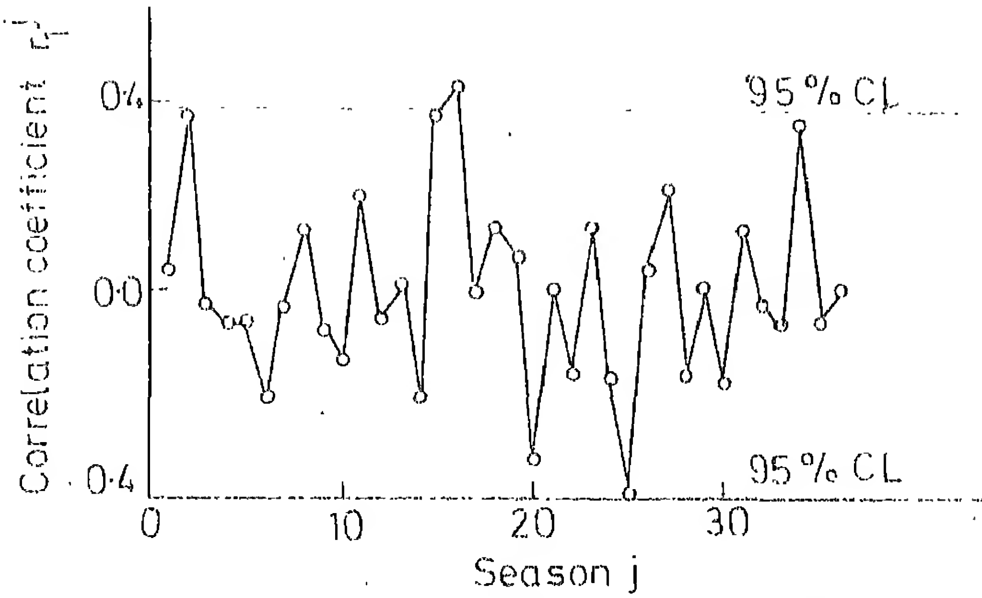


FIG.4.20 SEASONWISE CORRELATION OF TENDAILY RESIDUA (MLS) OF NONSTATIONARY FIRST ORDER AR MODEL

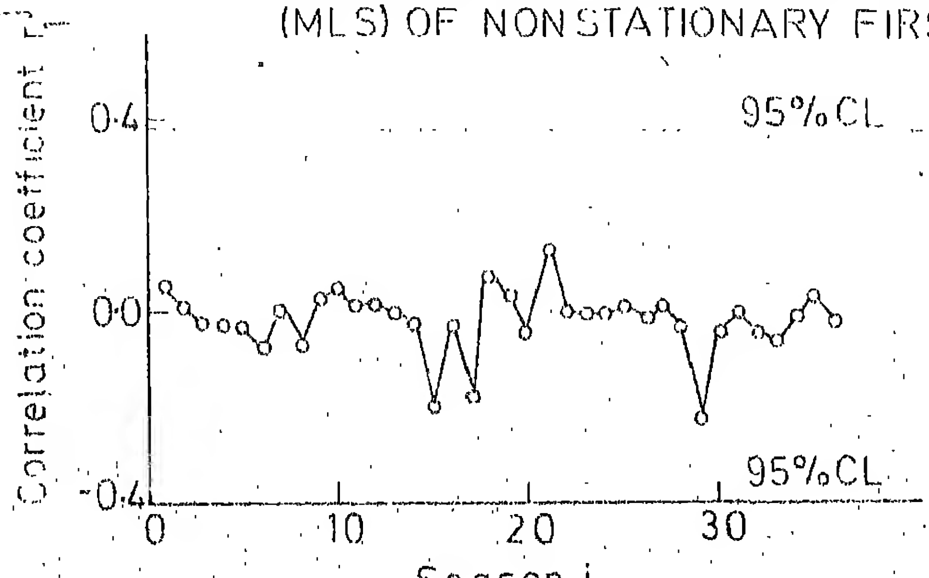


FIG.4.21 SEASONWISE CORRELATION OF TENDAILY RESIDUALS (MLS) OF NONSTATIONARY SECOND ORDER AR MODE

RIVER 1

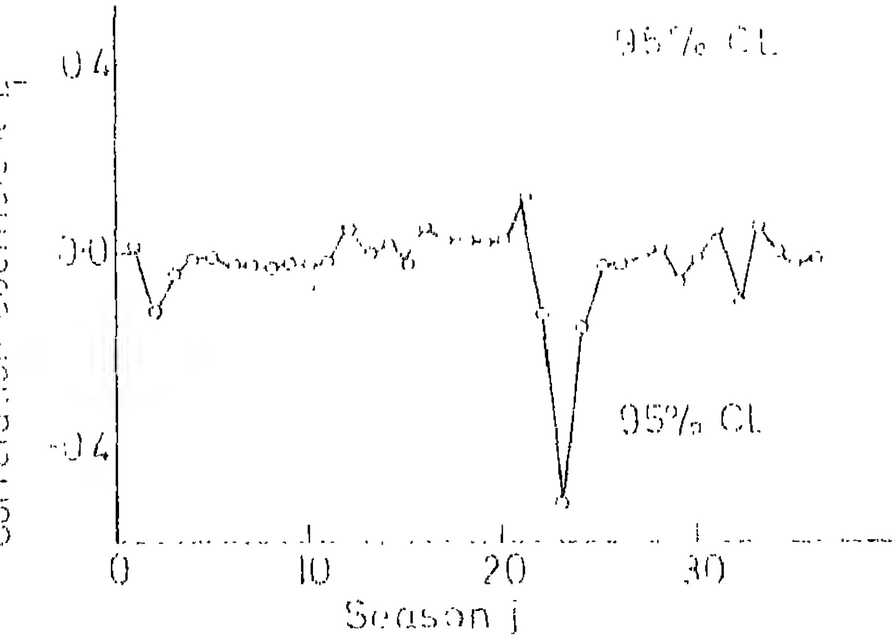


FIG. 4.22 SEASONWISE CORRELATION OF TENDAILY RESIDUALS (MLS) OF NONSTATIONARY SECOND ORDER AR MODEL, RIVER 2

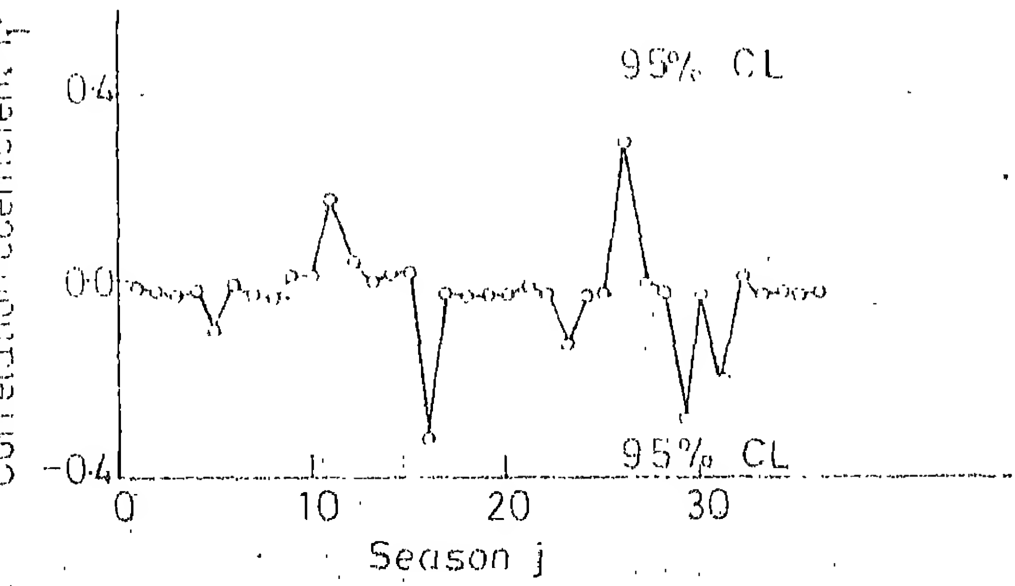
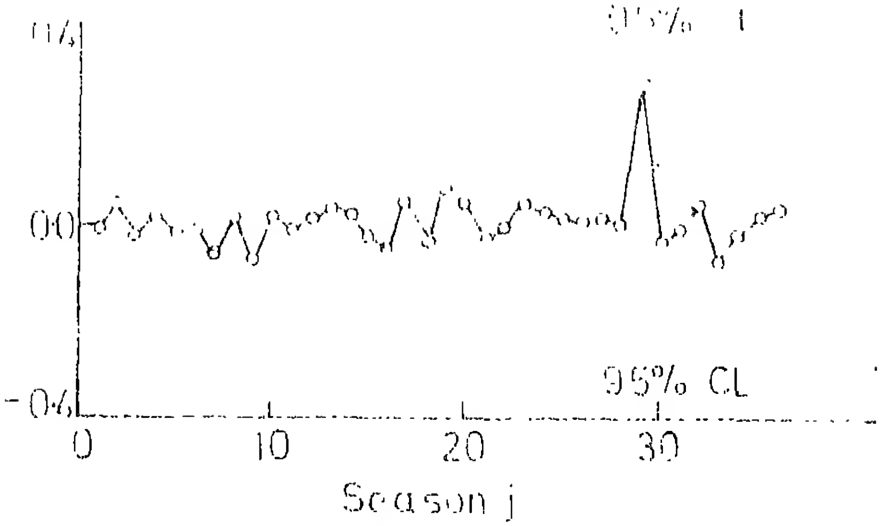
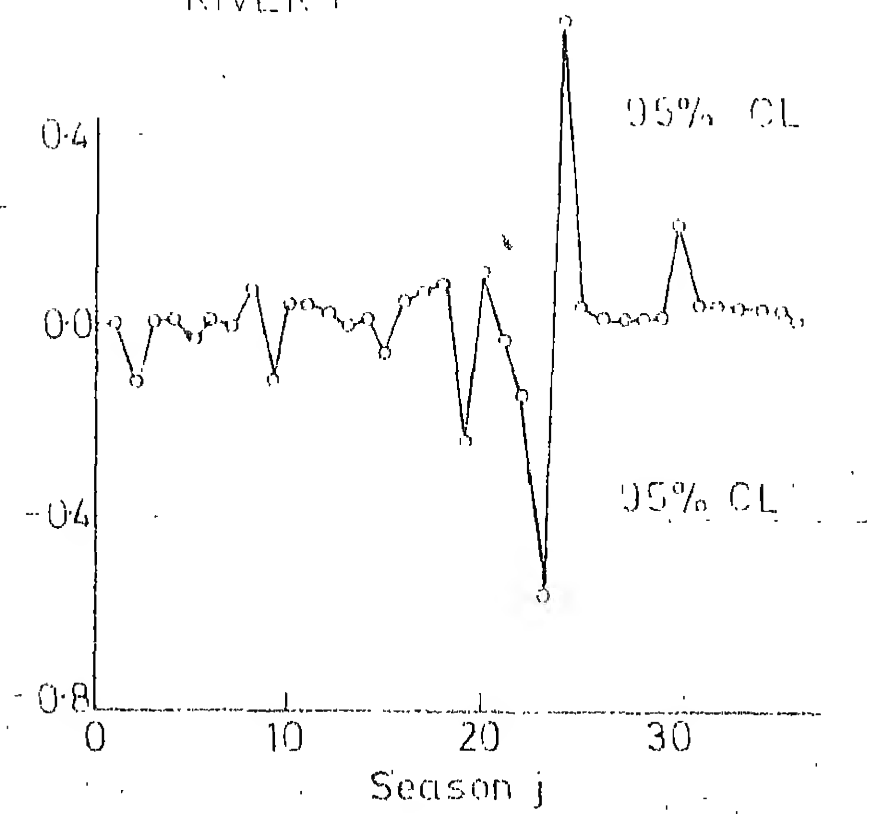


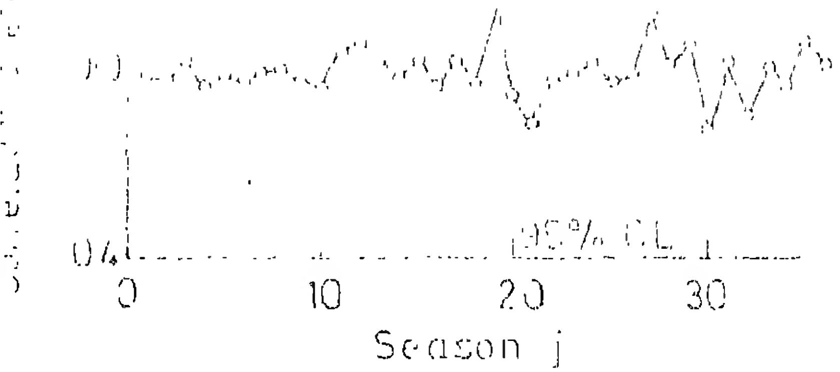
FIG. 4.23 SEASONWISE CORRELATION OF TENDAILY RESIDUALS (MLS) OF NONSTATIONARY SECOND ORDER MODEL, RIVER 3



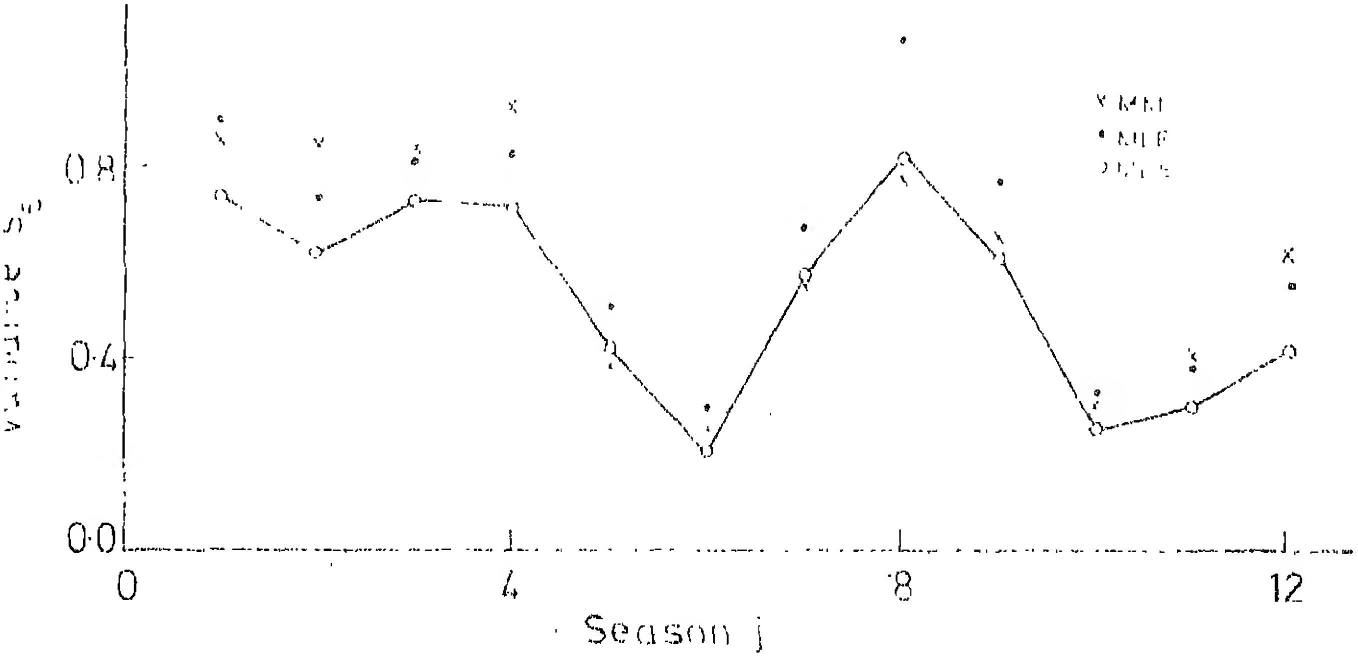
IG· 4·24 SEASONWISE CORRELATION OF TENDAILY RESIDUALS (MLS) OF NONSTATIONARY THIRD ORDER AR MODEL RIVER 1



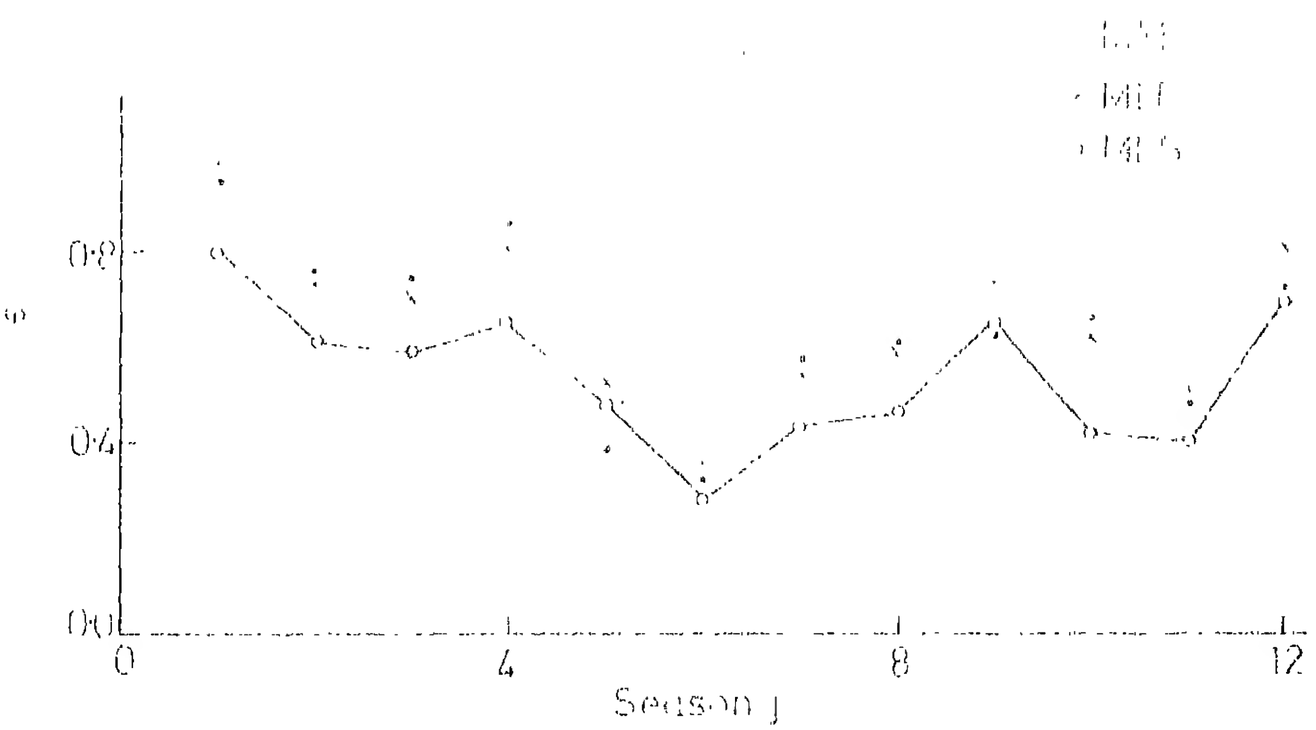
IG· 4·25 SEASONWISE CORRELATION OF TENDAILY RESIDUALS (MLS) OF NONSTATIONARY THIRD ORDER AR MODEL; RIVER 2



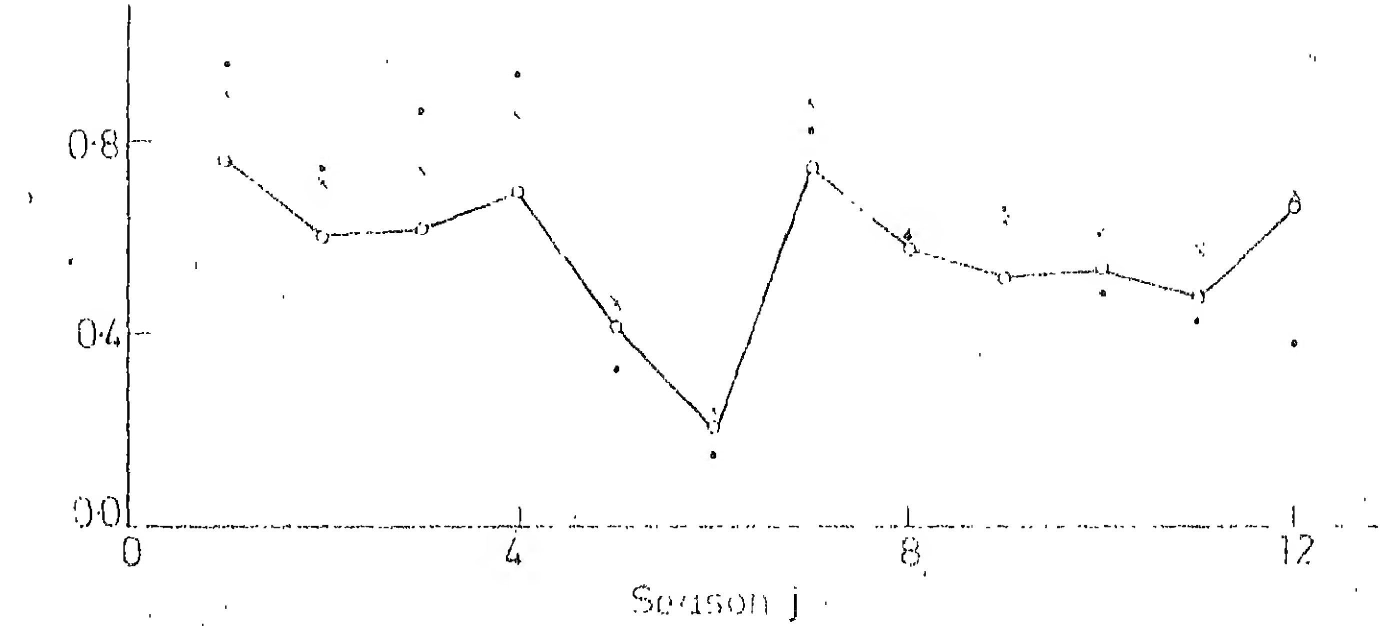
FIG·4·26 SEASONWISE CORRELATION OF TENDAILY RESIDUALS (MLS) OF NONSTATIONARY THIRD ORDER AR MODEL, RIVER 3



FIG·4·27 VARIANCE OF MONTHLY RESIDUALS OF NONSTATIONARY FIRST ORDER AR MODEL, RIVER 1



6·4·28 VARIANCE OF MONTHLY RESIDUALS OF NONSTATIONARY FIRST ORDER AR MODEL, RIVER 2



·4·29 VARIANCE OF MONTHLY RESIDUALS OF NONSTATIONARY FIRST ORDER AR MODEL, RIVER 3

the seasons of monsoon and winter precipitation that it is highest. This conclusion seems reasonable and is identical for all the three methods of standardisation.

(ii) Tendaily series: The variances of the residuals were obtained for the first, second and third order models using MM, MLE and MLS (Figs. 4.30 to 4.38). As in the case of monthly series, these variances were lowest during late autumn and early winter andhighest during the monsoon season. Also, it was noted that as the order of the model was increased from first to third, at each step there was some reduction in variance. Among the three methods of standardisation used, MLS again returned a lower residual variance on larger number of occasions than MM and MLE. For instance, for the third order model, when the coefficients were estimated by MLS there were less residual variances in 20 out of 36 seasons than when estimation was done by MM or MLE. It was also observed that MLS gave less variance more often during the monsoon months and spring. In autumn and winter, MM appeared to give less variance. A disturbing feature of MM was that, during some seasons, the residual variances were higher than the original variances.

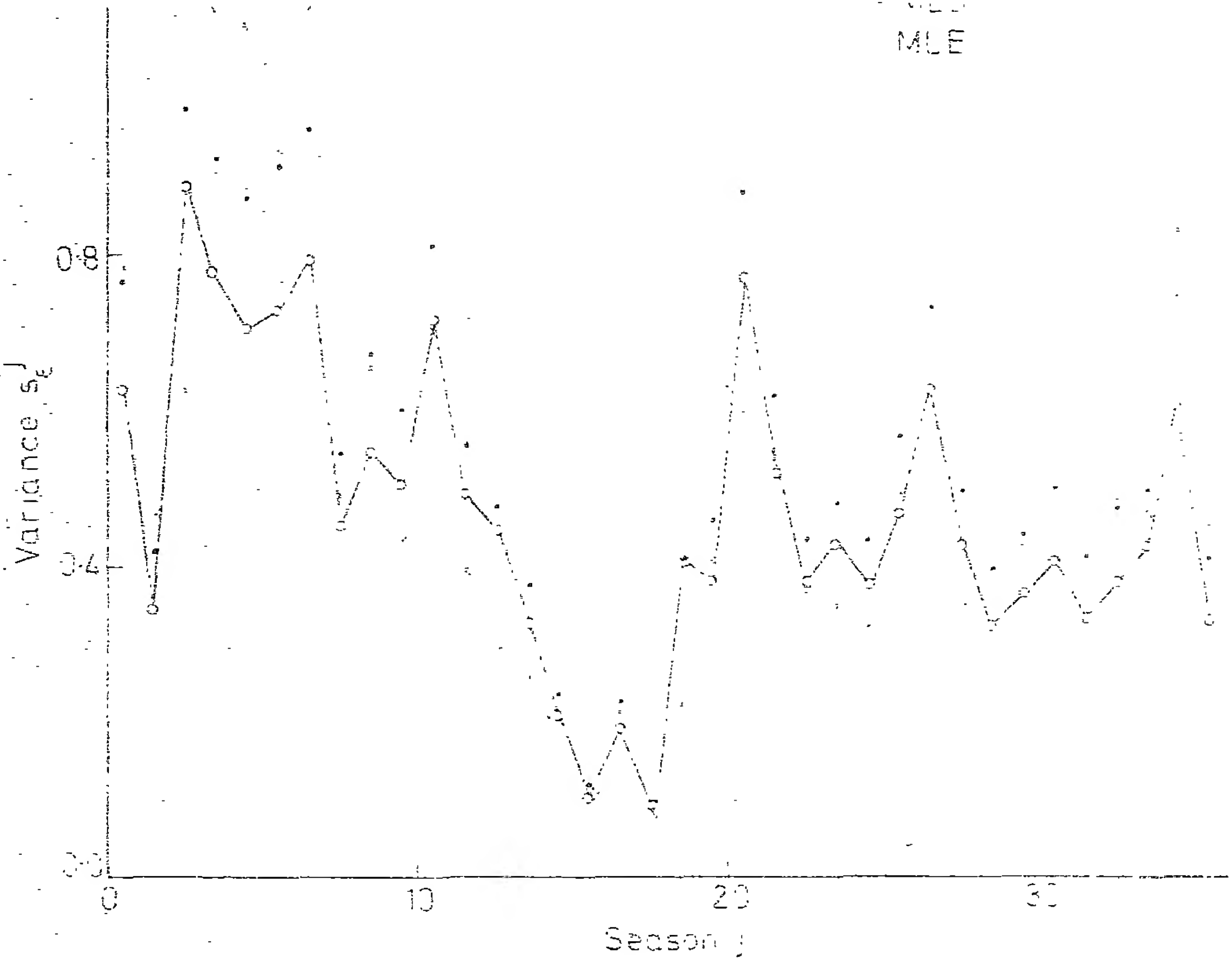
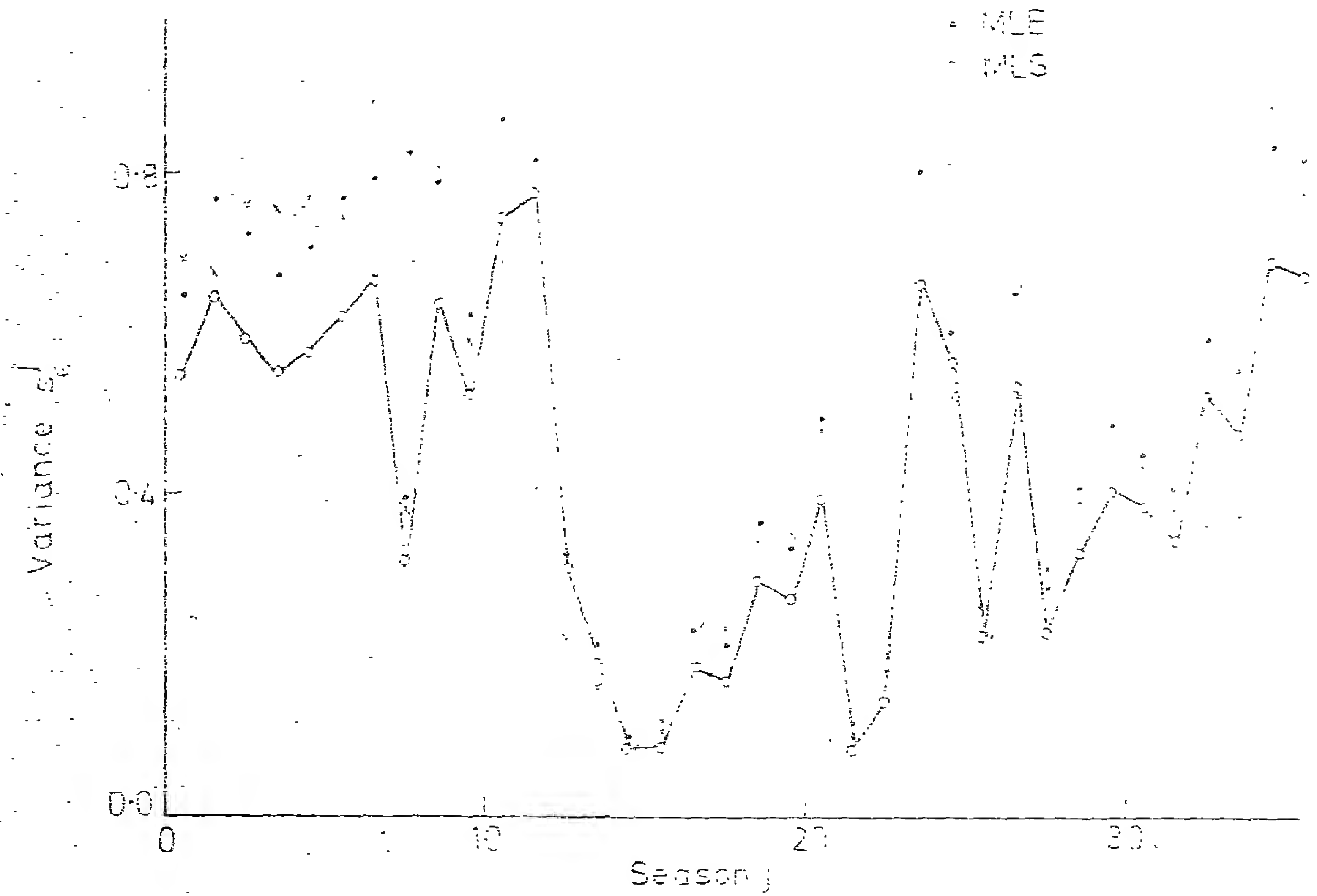


FIG. 4.30 VARIANCE OF TENDAILY RESIDUALS OF NONSTATIONARY FIRST ORDER AR MODEL, RIVER 1



FIG·4·31 VARIANCE OF TENDAILY RESIDUALS OF NONSTATIONARY FIRST ORDER AR MODEL, RIVER 2

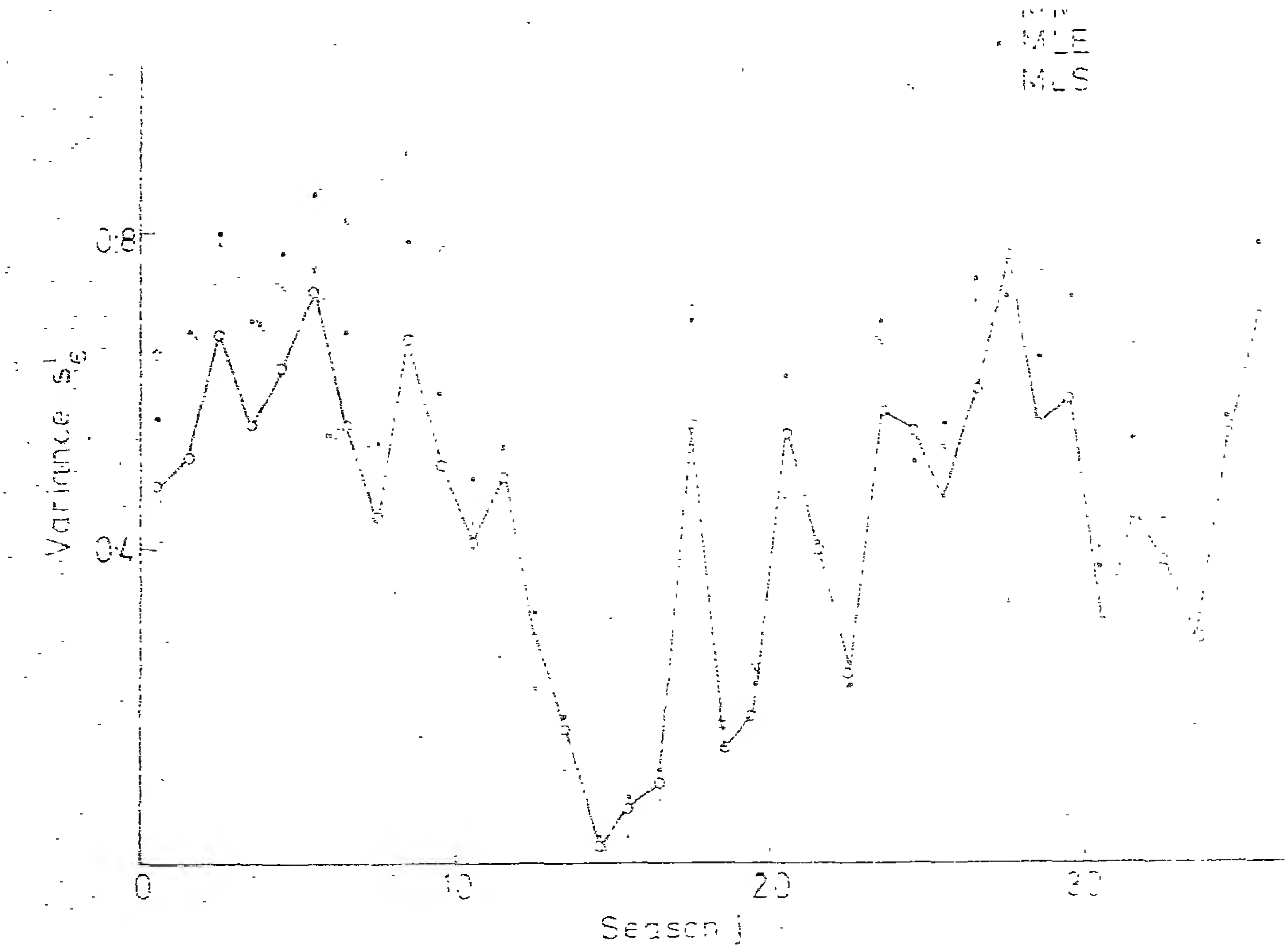


FIG 4.32 VARIANCE OF TENDAILY RESIDUALS OF NONSTATIONARY FIRST ORDER AR MODEL RIVER 3

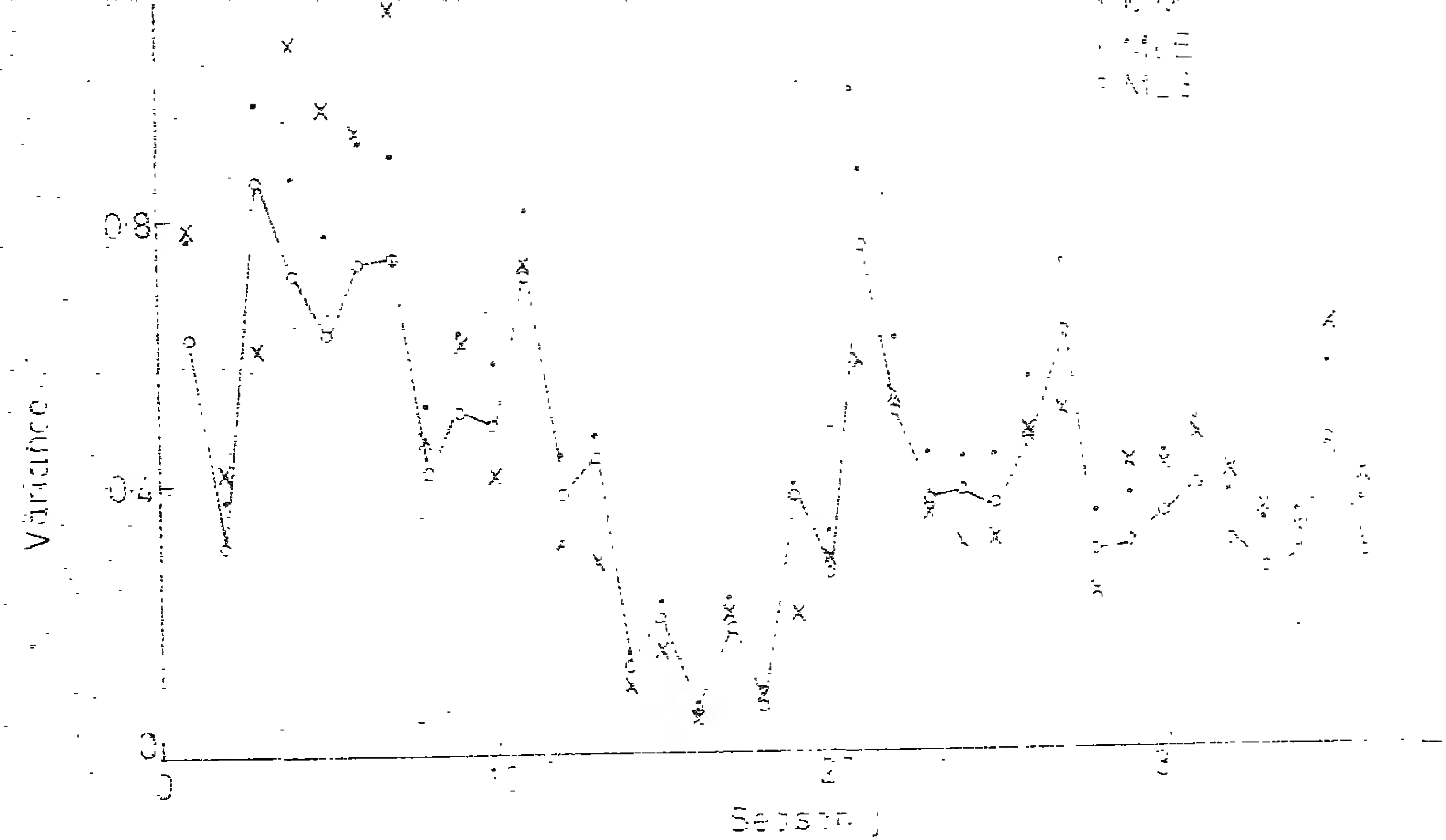


FIG.4.33 VARIANCE OF TENDAILY RESIDUALS OF NONSTATIONARY
SECOND ORDER AR MODEL, RIVER 1

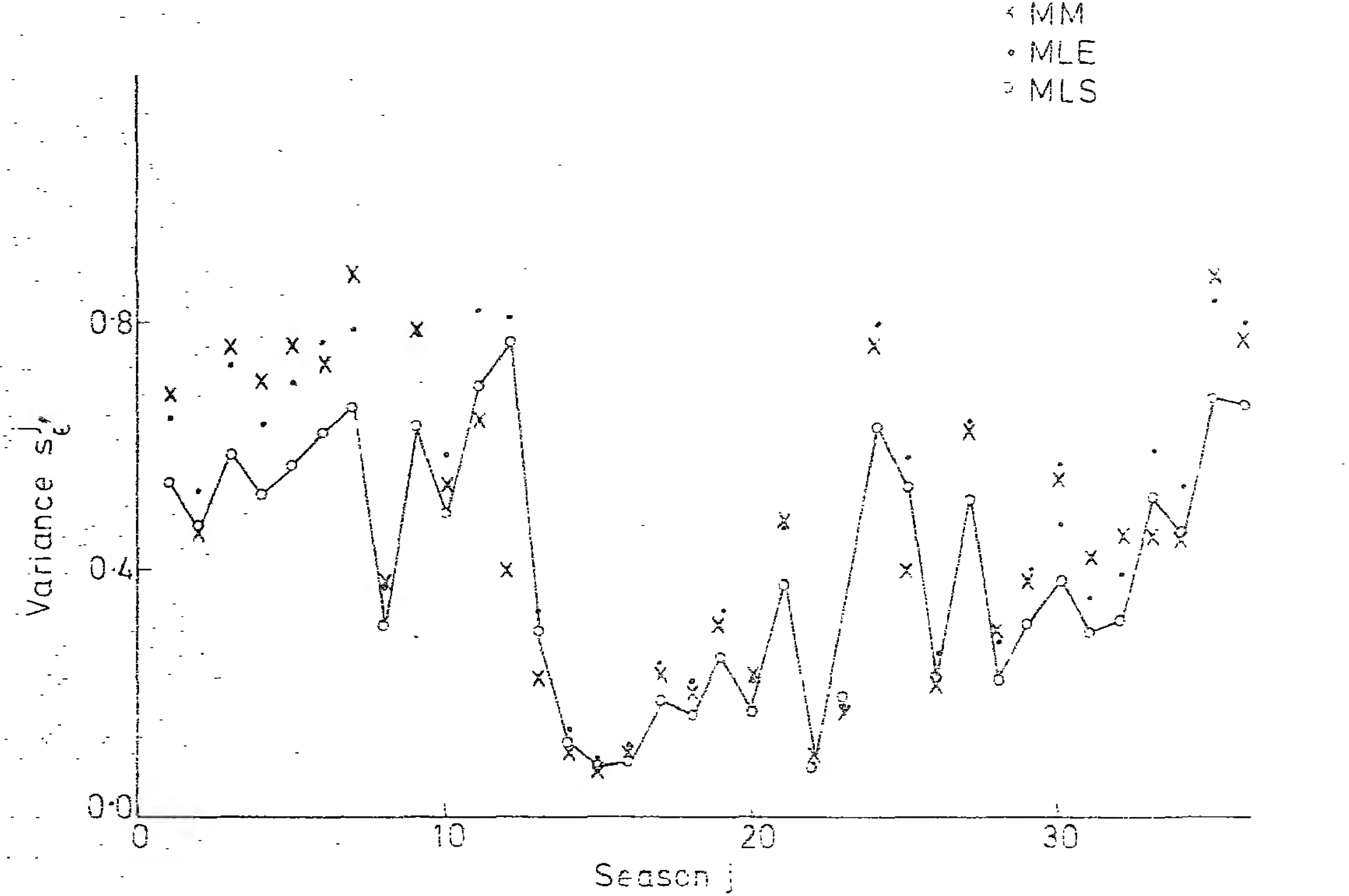


FIG. 4.34 VARIANCE OF TENDAILY RESIDUALS OF NONSTATIONARY SECOND ORDER AR MODEL, RIVER 2

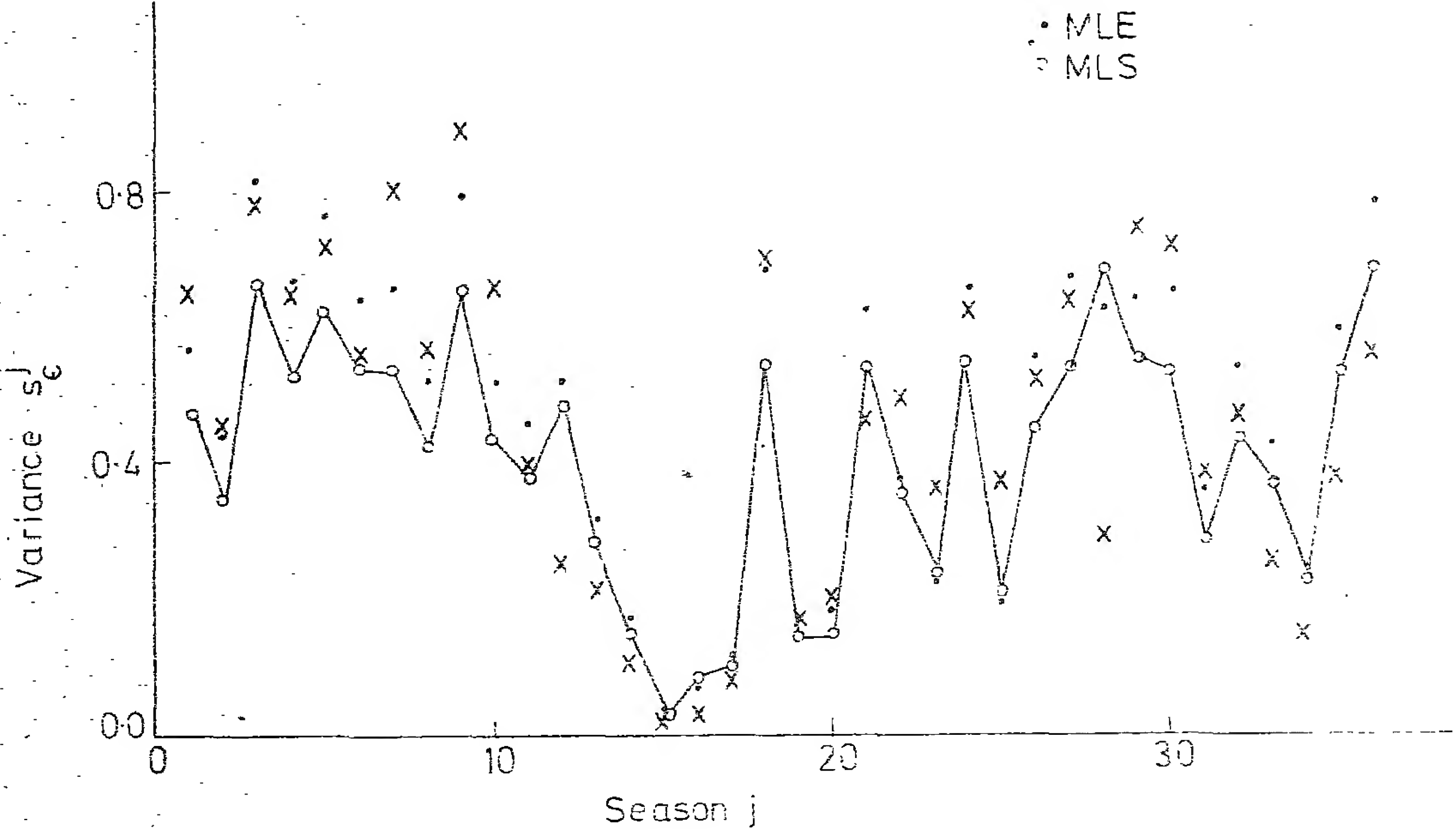


FIG. 4·35 VARIANCE OF TENDAILY RESIDUALS ON NONSTATIONARY SECOND ORDER AR MODEL, RIVER 3

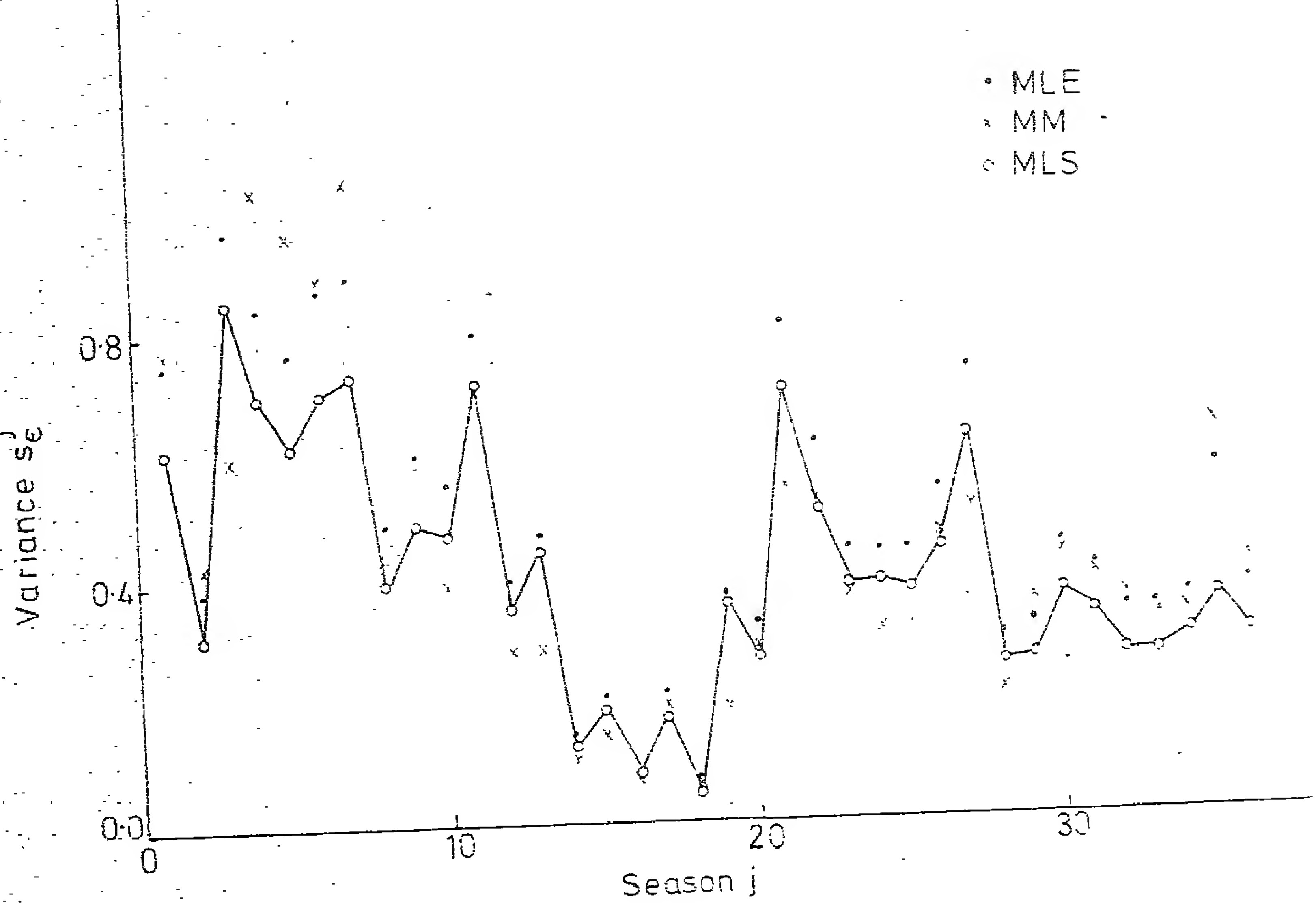
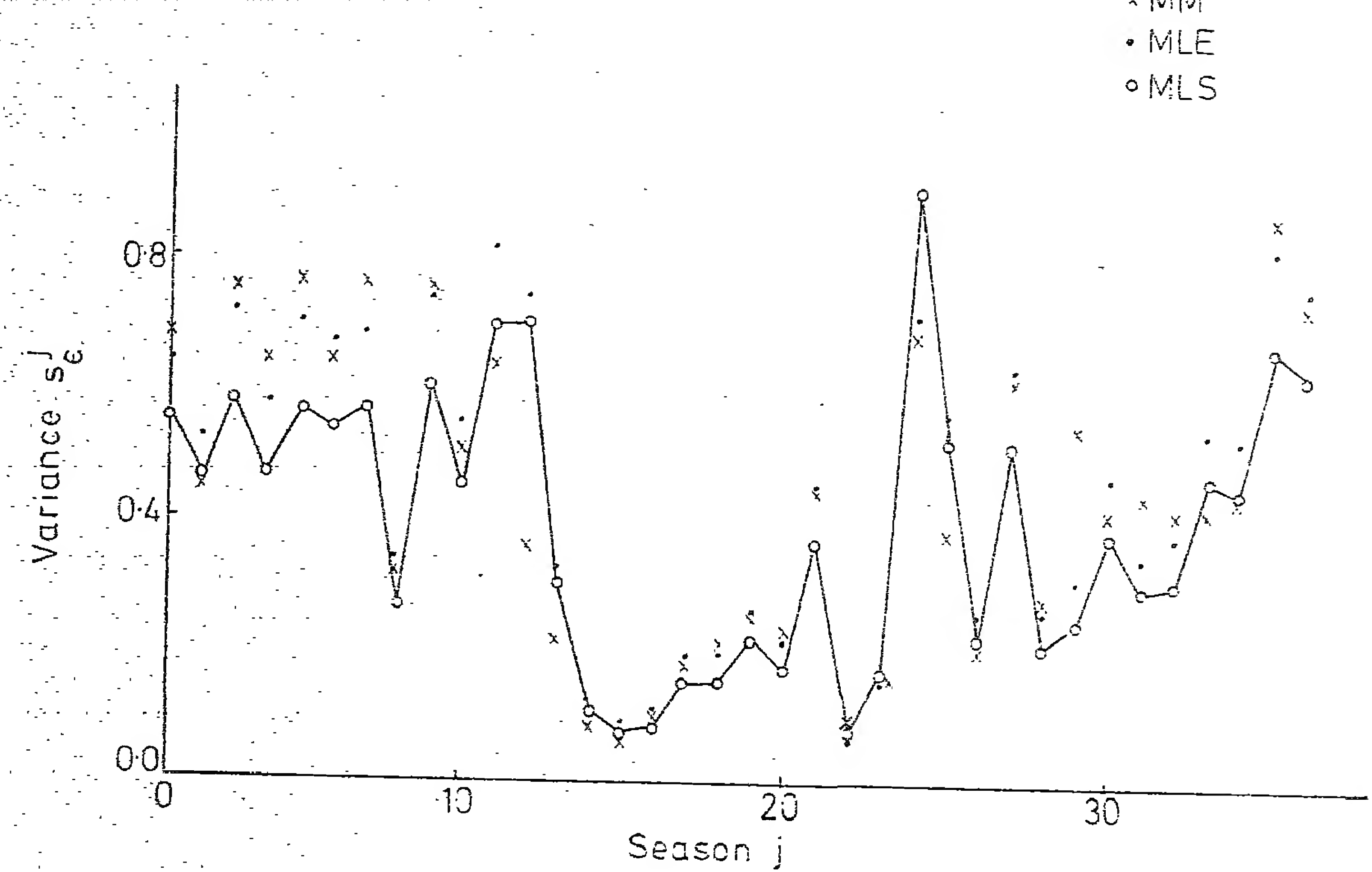


FIG.4.36 VARIANCE OF TENDAILY RESIDUALS OF NONSTATIONARY THIRD ORDER AR MODEL RIVER 1



FIG·4·37 VARIANCE OF TENDAILY RESIDUALS OF NONSTATIONARY THIRD ORDER AR MODEL, RIVER 2

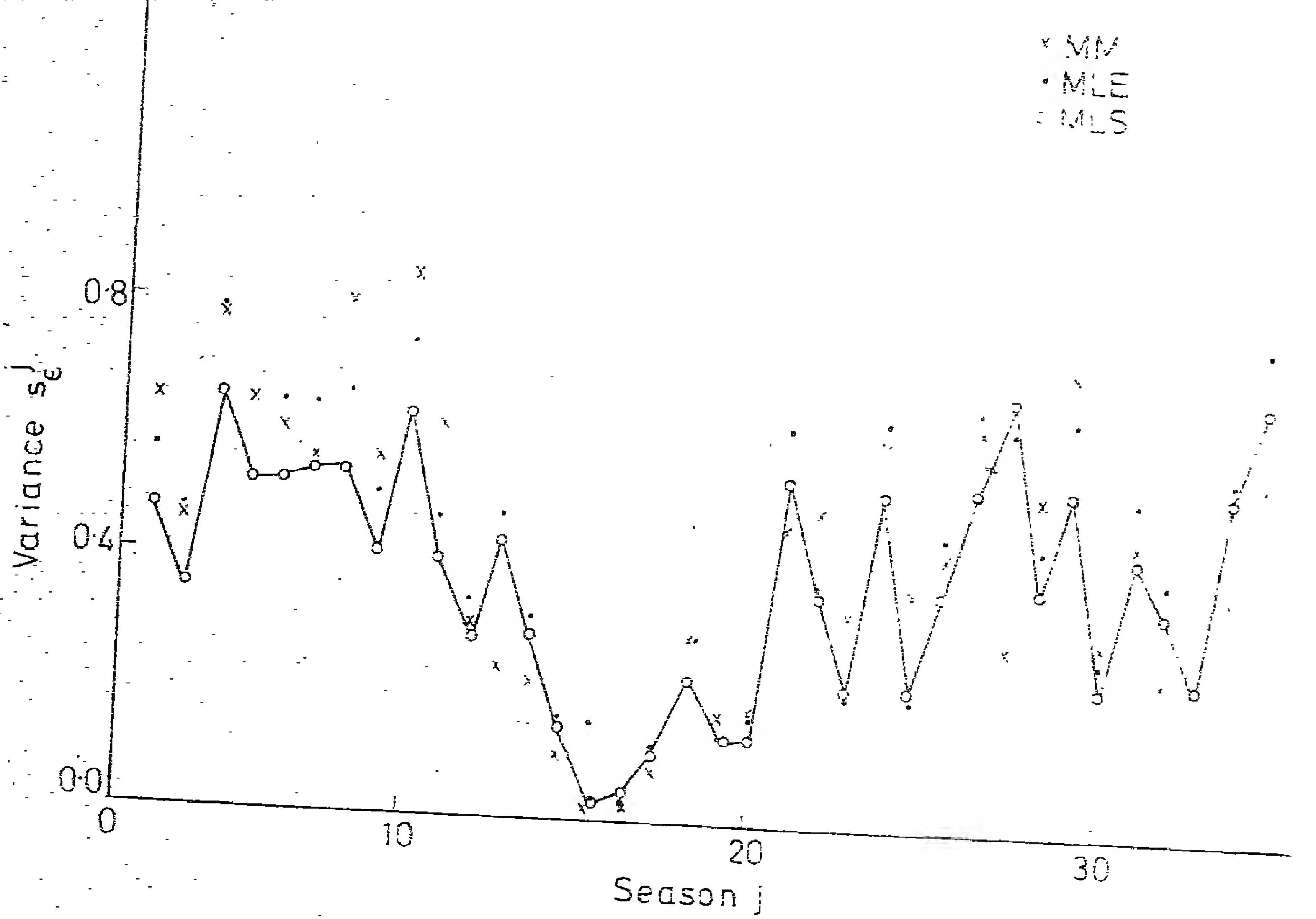


FIG. 4.38 VARIANCE OF TENDAILY RESIDUALS OF NONSTATIONARY THIRD ORDER AR MODEL RIVER 3

4.3 Conclusion

In selecting a univariate model for the logarithms of streamflows at each site, the following points may be noted:

(i) The residuals of the stationary model display significant seasonwise serial correlation. Increasing the order of the model does not show improvement in serial independence. On the other hand, the residuals of the nonstationary model are serially independent. This is because the persistence structure of the former is represented by unique autoregressive coefficient for the whole year, whereas the second model takes into consideration the time varying nature of the correlation structure. This no doubt increases the number of coefficients to be estimated.

(ii) Because of their stationary nature, the residuals of the stationary model have a unique value of the variance for the entire year. If the variances are estimated separately for each season using the above residuals, the residual variances are sometimes higher than the original variance. For the nonstationary model in which the parameters are estimated on a seasonal basis such a problem does not arise.

(iii) For the stationary models, the entire length of the historical data, after due transformation and standardisation,

is treated as a single sample and the coefficients are estimated. The nonstationary analysis divides the total sample into samples of smaller length and the coefficients are estimated from such small samples. Reduction in sample size results in a reduced reliability in the latter case. Likewise, when the residuals are checked for independence, the correlogram of residuals in a stationary model treats the entire data length as a single sample while the nonstationary model tests the correlation coefficients of residuals seasonally, utilising samples of shorter length. For the monthly series, the reduction is from 12 to 1 and for the tendaily series from 36 to 1.

(iv) The decoupling of the multivariate model implies that the univariate residuals should as far as possible be pure random. The use of a stationary AR model results in a nonstationary series of residuals with a seasonally varying correlation structure. On the other hand, the use of the nonstationary AR model results in a stationary residual series. Hence for the data considered, it is necessary to use a nonstationary AR model for representing the univariate process.

On the basis of the above considerations the following univariate models are adopted for further study: The annual series can be taken as serially uncorrelated. As such, a

zeroeth order univariate model seems suitable to represent this series. A first order nonstationary AR model is satisfactory for the univariate modelling of the monthly data. A third order nonstationary AR model is satisfactory for the univariate modelling of the tendaily data. For river 1, however, the representation seems to be not satisfactory with reference to the probability distribution. As normality is not an essential criterion, a nonstationary third order AR model is chosen for all the three rivers.

5. MULTIVARIATE MODELLING

5.1 Decoupled Multivariate Stochastic Models

The univariate residual series $\varepsilon(t)$ are determined from the nonstationary model

$$x(t) = \sum_{i=1}^{p} \phi_i^j \, x(t-i) + \varepsilon(t), \quad j = 1,2,\ldots s \quad (5.1)$$

where $\varepsilon(t)$ is a serially independent random variable with mean zero and variance $(s_\varepsilon^j)^2$, and s is the number of seasons within an year. When the number of stations is greater than one, the model can be represented as

$$x_k(t) = \sum_{i=1}^{p} \phi_i^j(k) \, x_k \, t-i) + \varepsilon_k(t), \quad k = 1,2,\ldots K \text{ and } j = 1,2,\ldots s \quad (5.2)$$

where $\varepsilon_k(t)$ is a serially independent random variable at time t with mean zero and variance $(s_{\varepsilon_k}^j)^2$ in season j for the k-th site and K is the total number of sites.

The set of serially uncorrelated series $\varepsilon_k(t)$ with $k = 1,\ldots K$ constitutes the multivariate time series in this study. Although decoupling of the univariate model has removed the internal dependence of each of the component time series, external dependence may still be present among the time series. Knowing the standard deviation $s_{\varepsilon_k}^j$, the set of $\varepsilon_k(t)$ series is standardised to an $\{E(t)\}$ series

with zero mean and unit standard deviation. Hence for the nonstationary univariate residuals

$$\{E(t)\} = \{e_1(t), \quad e_2(t), \ldots\ldots e_K(t)\}^T$$
$$= \left\{ \frac{\varepsilon_1(t)}{s^j_{\varepsilon_1}}, \frac{\varepsilon_2(t)}{s^j_{\varepsilon_2}}, \ldots\ldots \frac{\varepsilon_K(t)}{s^j_{\varepsilon_K}} \right\}^T \tag{5.3}$$

$\{E(t)\}$ constitutes the multivariate series of serially independent components. It may be represented by the multivariate ARMA model of order (p, q) as follows:

$$\{E(t)\} = \sum_{u=1}^{p} [C^j_u] \{E(t-u)\} + \sum_{v=1}^{q} [D^j_v] \{\eta(t+1-v)\},$$
$$j = 1,2,\ldots s \tag{5.4}$$

Earlier studies in multivariate modelling of streamflows (Matalas, 1967, Young and Pisano, 1968) have indicated that a multivariate first order autoregressive model may be sufficient for the representation of multisite streamflows. Hence in this study it was assumed that u = 1 and v = 1 so that the above equation reduces to the nonstationary multivariate first order autoregressive model.

$$\{E(t)\} = [C^j] \{E(t-1)\} + [D^j] \{\eta(t)\} \tag{5.5}$$

Assumption of stationarity results in $[C^j]$ and $[D^j]$ being independent of j, so that

$$\{E(t)\} = [C] \{E(t-1)\} + [D] \{\eta(t)\} \tag{5.6}$$

which is of the same form as proposed by Matalas (1967).

For the nonstationary model, the coefficient matrices $[C^j]$ and $[D^j]$ may be evaluated in a manner analogous to the procedures of Matalas (1967) and Young and Pisano (1968). Letting the lagzero and lagone correlation matrices among the $\{E(t)\}$ series for the j-th season to be represented by $[M_0^j]$ and $[M_{-1}^j]$ respectively, the coefficients of the model may be estimated by

$$[C^j] = [M_{-1}^j]\,[M_0^j]^{-1} \quad \text{and} \tag{5.7}$$

$$[D^j][D^j]^T = [M_0^j] - [C^j][M_{-1}^j]^T \tag{5.8}$$

Assuming a lower triangular form for the matrix $[D^j]$, a recursive method of obtaining a unique solution for $[D^j]$ was presented by Young and Pisano (1968) for the case $j = 1$. The same procedure was used in this study for estimating the coefficients. In case the data are nonconcurrent and $[M_0^j]$ is not positive definite, the procedures suggested by Fiering (1968); Crosby and Maddock (1970) and Fuller (1974) may be used to adjust the elements of the correlation matrix in order to make them consistent. They were not required in this study for the stationary model as concurrent data were used.

5.2 Decoupled Multivariate Stationary Model

5.2.1 Estimation of parameters

The annual series are serially independent and pure random. So they constitute the standardised multivariate data. Three different univariate models have been fitted to the monthly and tendaily series using MM, MLE and MLS for parameter estimation. The univariate standardised residuals corresponding to the three methods were determined for fitting multivariate models. When the results for the three methods of standardisation are comparable, the results for only one method are presented. In the case of all the series, a stationary multivariate first order AR model was initially assumed. The coefficient matrices [C] and [D] were estimated using the procedure in Sec.5.1.

5.2.2 Significance tests on parameters

Matalas and Wallis (1972) have pointed out that $[M_0]$ and $[DD^T]$ should be positive definite in order that the multivariate first order AR model may be fitted to any data series. Let $|[\cdot]|$ or $|\cdot|$ define the determinant of $[\cdot]$. Thus even if $|M_0| > 0$, it should be separately checked if $|DD^T| > 0$. It was found that for stationary multivariate models, both the above requirements were met with for annual, monthly and tendaily series.

[C] matrix is a measure of the lagone crosscorrelation among the $\{E(t)\}$ series. Whether it is significantly different from zero may be tested by a multidimensional test of significance due to Anderson (1958). Let

$$F_o = |M_o| \quad \text{and} \tag{5.9}$$

$$F_1 = |M_o| - |[C][M^T_{-1}]| \tag{5.10}$$

Define a statistic Z given by

$$Z = \frac{F_o - F_1}{F_o} \cdot \frac{N - m}{m} \tag{5.11}$$

where N is the size of the sample, m is the order of the matrix and m = K. Z is approximately distributed as chi-square with m degrees of freedom. If the estimated Z is less than the theoretical value, at, say, 95% level, then the estimated Z statistic and hence [C] are not statistically significant at that level.

In physical systems, there may be physical reasons to justify the presence of serial or crosscorrelation even though they may not be statistically significant. It is the general practice to take into account these correlations when they explain more than, say, 5% of the variance of the series even if the coefficients are not statistically significant. A comparison of F_1 (Eq. 5.10) with $[DD^T]$ (Eq. 5.8) indicates that F_o is a measure of the multivariate variance while F_1 is

a measure of the residual multivariate variance. Since the rivers considered are tributaries to the same major river and they lie adjacent to each other in an essentially homogeneous hydrometeorological region, a correlation seems meaningful even when they are not statistically significant provided they explain more than 5% of the variance.

Annual series: The results of the analyses for the annual series are given in Table 5.1. $[M_0]$ and $[DD^T]$ are both positive definite and the Z statistic is not significant at 95% level. However, the variance explained by [C] is of the order of 34% of the total variance and appears to be significant. Hence [D] seems to be physically meaningful even though it is not statistically significant and so it is retained in the model.

Monthly series: The results of the analyses for the monthly series are given in Table 5.2. $[M_0]$ and $[DD^T]$ are both positive definite in all the cases. The values of Z statistic are not significant at 95% level implying thereby that the [C] matrices are not statistically significant at that level. The variances explained by the [C] matrices are of the order of 4% of the total variance and hence the [C] matrices are considered to be not physically significant. Hence the monthly series can be represented by a multivariate model of zeroeth order, viz.,

TABLE 5.1 PARAMETER ESTIMATES FOR STATIONARY MULTIVARIATE MODEL (ANNUAL SERIES)

$$[C] = \begin{bmatrix} -.62 & -.43 & .95 \\ -.44 & -.42 & .64 \\ -.28 & -.53 & .60 \end{bmatrix} \qquad [D] = \begin{bmatrix} .89 & & \\ .62 & .70 & \\ .74 & .39 & .42 \end{bmatrix}$$

$$Z_E = 4.00; \; Z_{T,95\%} = 7.82 \qquad |DD^T| = .07$$

$$\frac{F_o - F_1}{F_1} \simeq 0.5$$

TABLE 5.2 PARAMETER ESTIMATES FOR STATIONARY FIRST ORDER MULTIVARIATE MODEL (MONTHLY SERIES)

$$\begin{array}{cccc} & \text{MM} & \text{MLE} & \text{MLS} \\ [C] = & \begin{bmatrix} .03 & -.14 & .03 \\ .13 & -.06 & -.02 \\ .16 & -.08 & -.03 \end{bmatrix} & \begin{bmatrix} .03 & -.10 & -.02 \\ .14 & -.05 & -.01 \\ .15 & -.08 & -.03 \end{bmatrix} & \begin{bmatrix} .05 & -.13 & .02 \\ .12 & -.06 & -.01 \\ .17 & -.10 & -.02 \end{bmatrix} \\ & Z_E = 3.96; & Z_E = 4.20; & Z_E = 3.88; \\ & Z_{T,95\%} = 7.82 & Z_{T,95\%} = 7.82 & Z_{T,95\%} = 7.82 \\ [D] = & \begin{bmatrix} .99 & & \\ .58 & .81 & \\ .42 & .37 & .82 \end{bmatrix} & \begin{bmatrix} 1.00 & & \\ .62 & .78 & \\ .49 & .32 & .81 \end{bmatrix} & \begin{bmatrix} .99 & & \\ .58 & .81 & \\ .35 & .31 & .88 \end{bmatrix} \\ & |DD^T| = .42 & |DD^T| = 40 & |DD^T| = .50 \end{array}$$

$$\frac{F_o - F_1}{F_1} \simeq 0.04$$

$$\{E(t)\} = [D] \{\eta(t)\} \tag{5.12}$$

This result is similar to that of Yevjevich for the net basin supplies to the Great Lakes of North America (1975). Assuming [D] to be lower diagonal, the elements of the [D] matrix are estimated from the equation

$$[DD^T] = [M_o] \tag{5.13}$$

The elements of the [D] matrix estimated from Eqs. 5.12 and 5.13 are given in Table 5.2(a). A comparison of Tables 5.2 and 5.2(a) indicates that the coefficients of the [D] matrix have increased only very slightly from the first order model to the zeroth order model. This confirms the adequacy of the zeroeth order model.

TABLE 5.2(a) PARAMETER ESTIMATES FOR STATIONARY ZEROETH ORDER MULTIVARIATE MODEL (MONTHLY SERIES)

$$\begin{array}{cccc} & \text{MM} & \text{MLE} & \text{MLS} \\ [D] = & \begin{bmatrix} 1.00 & & \\ .59 & .82 & \\ .44 & .38 & .84 \end{bmatrix} & \begin{bmatrix} 1.00 & & \\ .64 & .79 & \\ .51 & .34 & .83 \end{bmatrix} & \begin{bmatrix} 1.00 & & \\ .59 & .81 & \\ .37 & .33 & .89 \end{bmatrix} \\ & |DD^T| = .45 & |DD^T| = .43 & |DD^T| = .52 \end{array}$$

Tendaily sèries: The results of the analyses for the tendaily series are given in Table 5.3. $[M_o]$ and $[DD^T]$ are positive definite in all cases and the Z statistic is not significant at 95% level. The variances explained by [C] matrices are of the order of 2% of the total variance.

TABLE 5.3 PARAMETER ESTIMATES FOR STATIONARY FIRST ORDER MULTIVARIATE MODEL (TENDAILY SERIES)

$$
\begin{array}{cccc}
 & \text{MM} & \text{MLE} & \text{MLS} \\
[C] = & \begin{bmatrix} -.03 & .06 & -.02 \\ .06 & -.01 & -.06 \\ .01 & .11 & -.04 \end{bmatrix} & \begin{bmatrix} -.03 & .06 & -.02 \\ .06 & .00 & -.06 \\ .01 & .11 & -.05 \end{bmatrix} & \begin{bmatrix} -.04 & .07 & -.02 \\ .05 & -.01 & -.07 \\ .01 & .10 & -.05 \end{bmatrix} \\
 & Z_E = 5.53 & Z_E = 5.40 & Z_E = 5.93 \\
 & Z_T^{+} = 7.82 & Z_T^{+} = 7.82 & Z_T^{+} = 7.82 \\
[D] = & \begin{bmatrix} 1.00 & & \\ .56 & .83 & \\ .43 & .27 & .86 \end{bmatrix} & \begin{bmatrix} 1.00 & & \\ .56 & .83 & \\ .43 & .27 & .86 \end{bmatrix} & \begin{bmatrix} 1.00 & & \\ .54 & .84 & \\ .43 & .26 & .86 \end{bmatrix} \\
 & |DD^T| = .50 & |DD^T| = .50 & |DD^T| = .52
\end{array}
$$

$$\frac{F_o - F_1}{F_1} \simeq 0.02$$

Z_T^{+} is the theoretical value at 95% CL with K Degrees of freedom.

Hence the [C] matrices are considered to be statistically and physically not significant. The tendaily series can hence be represented by the multivariate zeroeth order AR model. The elements of [D] estimated from Eqs. 5.12 and 5.13 are given in Table 5.3(a). A comparison of the results of Tables 5.3 and 5.3(a) indicates that the corresponding elements differ only very slightly thus confirming the adequacy of the zeroeth order model.

TABLE 5.3(a) PARAMETER ESTIMATES FOR STATIONARY ZEROETH ORDER MULTIVARIATE MODEL (TENDAILY SERIES)

$$[D] = \begin{bmatrix} 1.00 & & \\ .57 & .83 & \\ .44 & .28 & .87 \end{bmatrix} \quad \begin{bmatrix} 1.00 & & \\ .57 & .83 & \\ .44 & .28 & .87 \end{bmatrix} \quad \begin{bmatrix} 1.00 & & \\ .55 & .85 & \\ .44 & .27 & .87 \end{bmatrix}$$

$$|DD^T| = .51 \qquad |DD^T| = .51 \qquad |DD^T| = .55$$

5.2.3 Testing of residuals

The multivariate residuals $\{\eta(t)\}$ were determined from the multivariate time series and the fitted models. They were then tested for normality by the chisquare test; independence within the time series by correlogram and spectral analyses; seasonwise independence within the time series by the seasonwise first serial correlation coefficient; independence among the time series by cross correlogram, and coherence spectra; and seasonwise independence among the time series by the seasonwise lagzero crosscorrelation coefficients.

Normality: The multivariate residuals for the annual series were tested for normality by the chisquare test (Table 5.4) and were accepted at 95% level. The results for the chisquare tests on the monthly series are given in Table 5.5. A comparison of these with those forthe univariate monthly residuals (Table 4.7) indicates that the results are comparable for MM and MLE and are better for the multivariate

TABLE 5.4 χ^2 ESTIMATES OF MULTIVARIATE ANNUAL RESIDUALS

RIVER	1	2	3
$\chi^2_{ESTIMATE}$	.2	1.0	.7
$\chi^2_{THEORETICAL^+}$	7.8	7.8	7.8

\+ Degrees of freedom = 3; Significance level = 95 %

TABLE 5.5 χ^2 ESTIMATES OF MULTIVARIATE MONTHLY RESIDUALS (STATIONARY ZEROETH ORDER MODEL)

NUMBER OF MONTHS WHEN HYPOTHESIS IS REJECTED AT 95% LEVEL

RIVER	STANDARDISATION BY		
	MM	MLE	MLS
1	2	2	1
2	1	1	1
3	1	1	0

residuals for MLS. The results of the chisquare test on the multivariate tendaily residual series are given in Table 5.6. A comparison with the results for univariate residuals (Table 4.7) indicates a better fit for station 2 and a comparable fit for stations 1 and 3 in the case of MM and MLE and a comparable fit in the case of MLS. The goodness of fit is hence considered satisfactory in the case of annual monthly and tendaily series.

Independence within the time series: The correlogram and power spectra of each of the time series in the multivariate residuals $\{\eta(t)\}$ are determined by standard procedures and tested for serial independence (Subsec. 4.1.3). The correlograms and power spectra for the annual series are shown along with the 95% confidence limits for the white noise series in Fig. 5.1(a) and Fig. 5.1(b). They indicate that the series are not significantly different from pure random series at 95% level. The correlograms and power spectra of the multivariate residuals for the monthly and tendaily series (MLS only) are shown in Figs. 5.2(a) and 5.2(b) and 5.3(a) and 5.3(b) respectively. The residuals are not significantly different from the white noise series at 95% level.

Seasonwise independence within the time series: The first serial correlation coefficients between consecutive seasons are estimated for each of the monthly and tendaily series and tested for significance. The results (for MLS only) are shown in Fig. 5.4(a) for the monthly series and Figure 5.4 (b) for the tendaily series. They indicate that the seasonwise serial correlation coefficients are statistically not significant at 95% level.

Independence among the time series: In order to test if the $\{\eta(t)\}$ residuals for each of the time series were mutually independent, the cross corrologram between

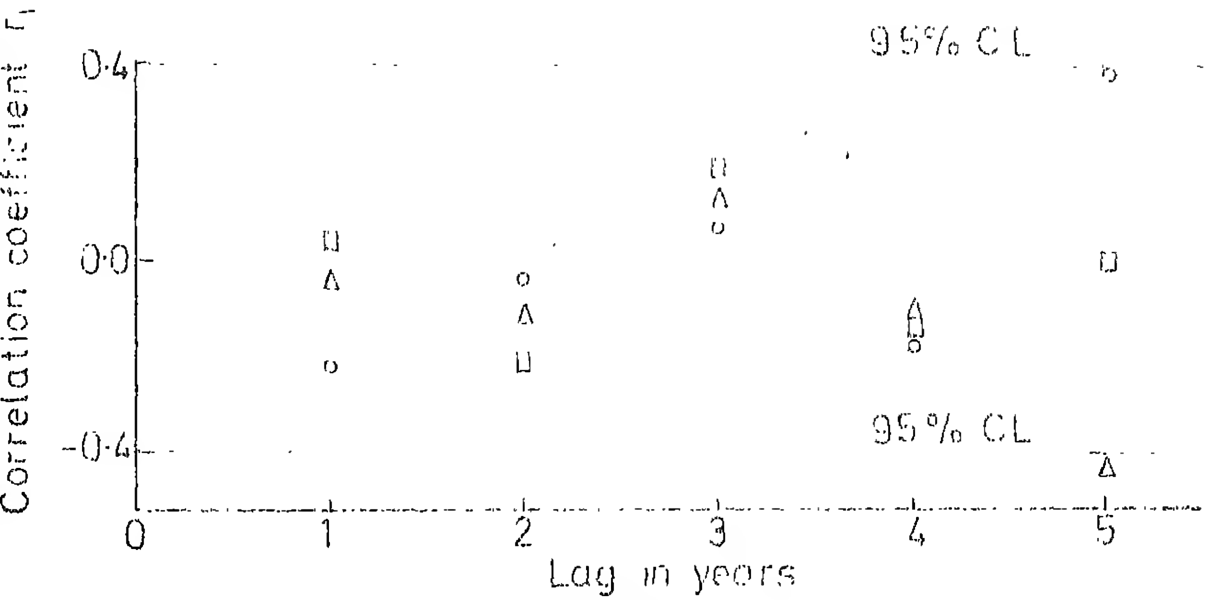
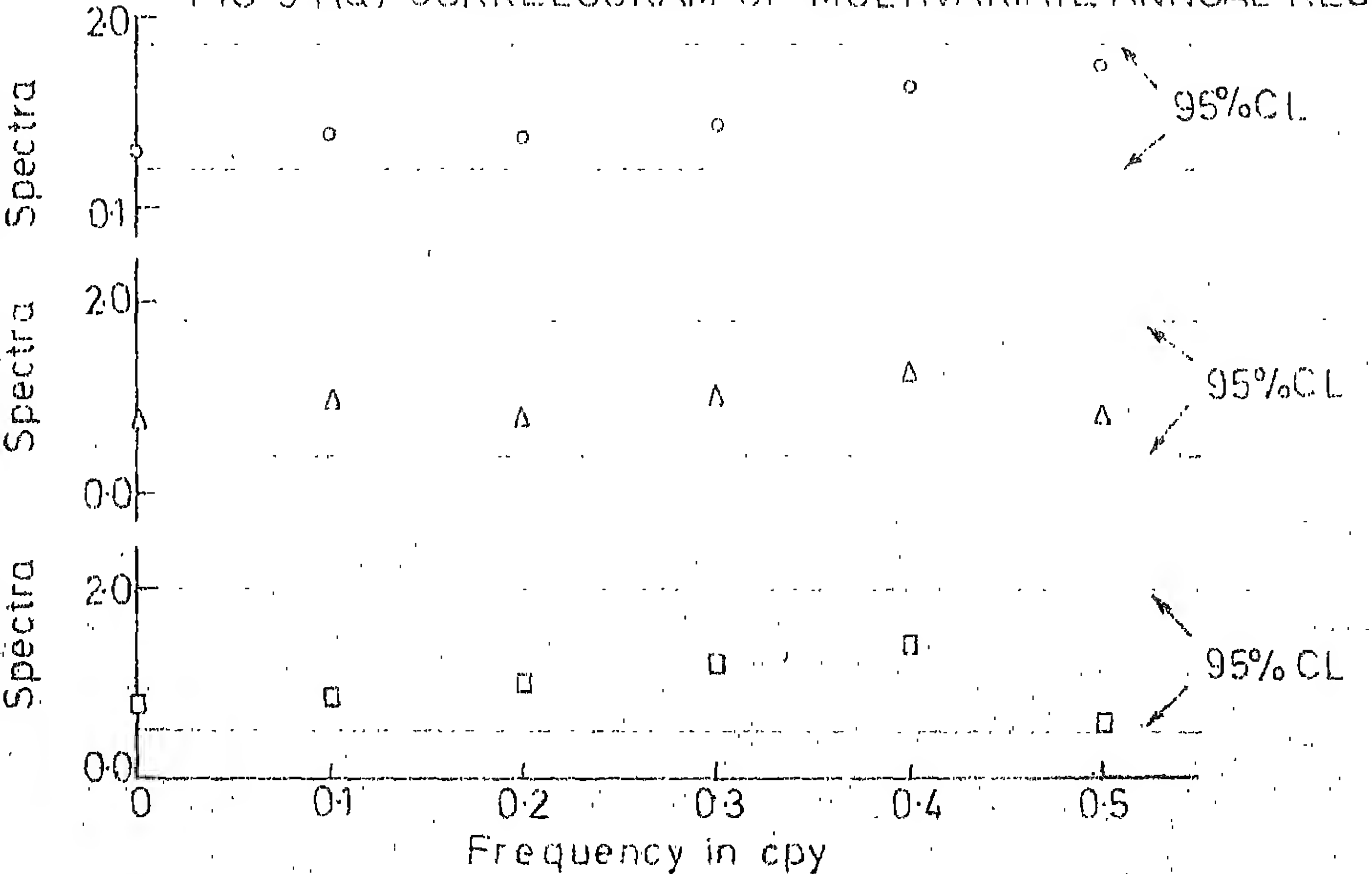


FIG 5·1(a) CORRELOGRAM OF MULTIVARIATE ANNUAL RESIDUA



FIG·5·1(b) POWER SPECTRA OF MULTIVARIATE ANNUAL RESIDUALS

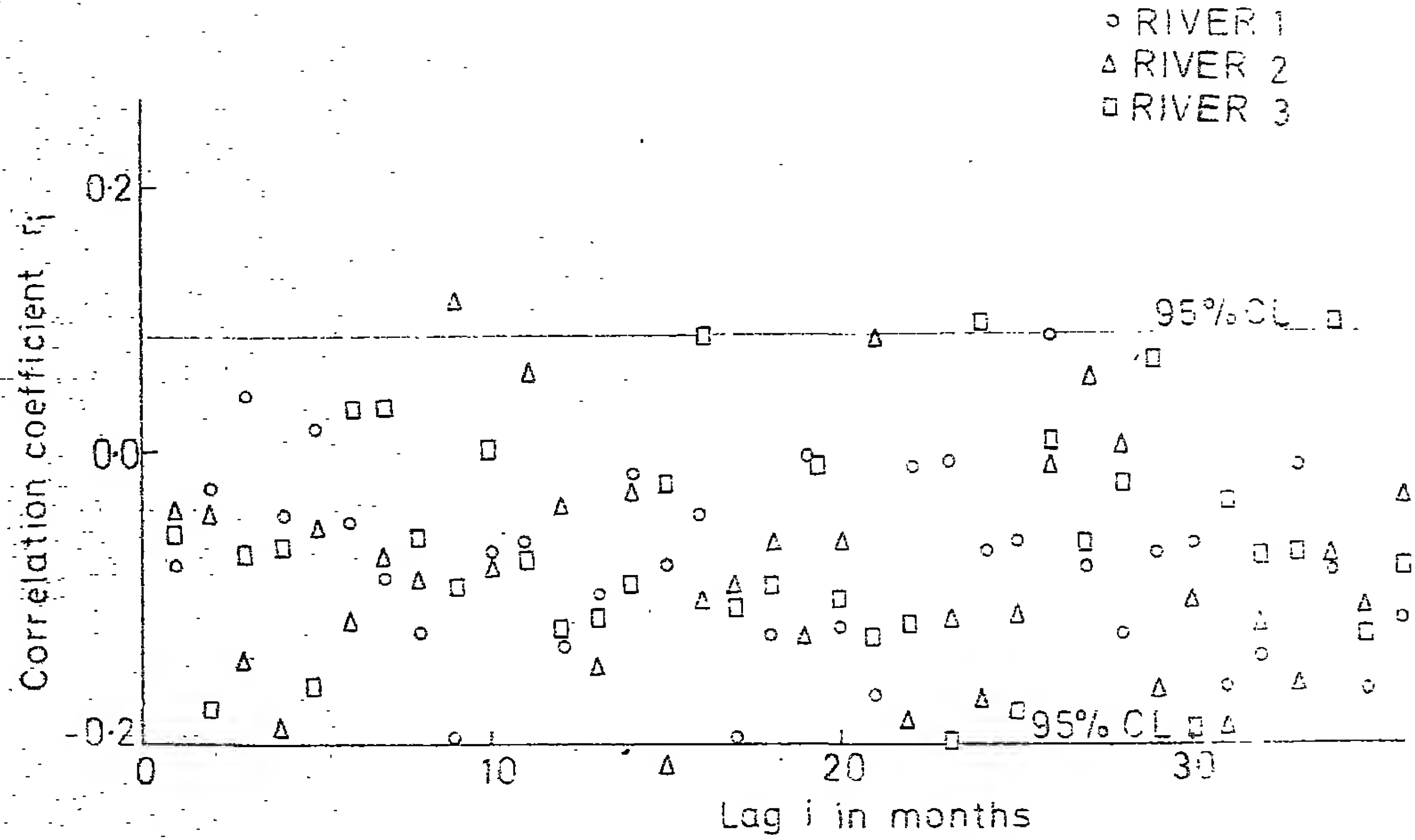


FIG. 5·2 (a) CORRELOGRAM OF MULTIVARIATE MONTHLY RESIDUALS (MLS) (STATIONARY MODEL)

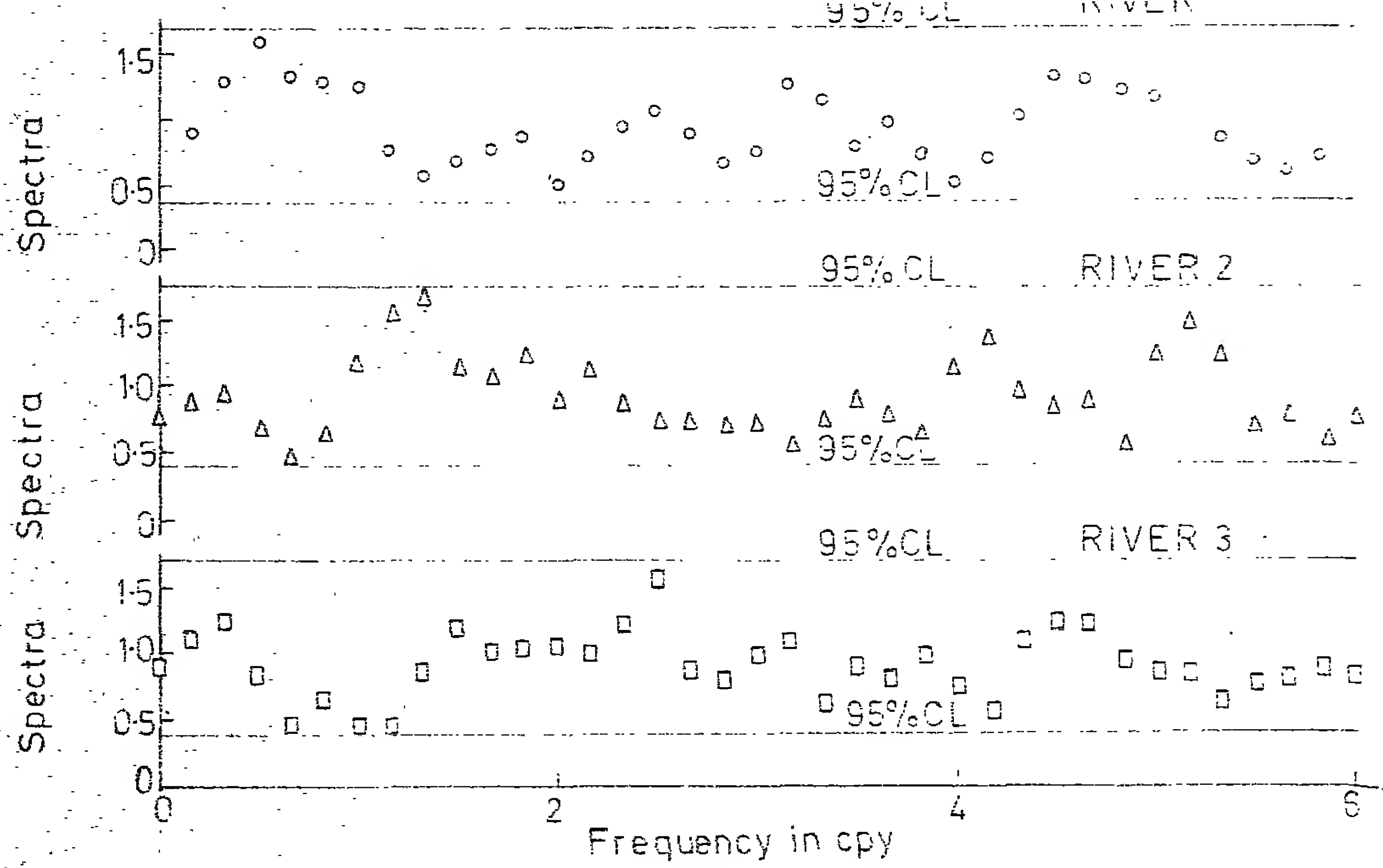


FIG. 5·2 (b) POWER SPECTRA OF MULTIVARIATE MONTHLY RESIDUALS (MLS) STATIONARY MODEL

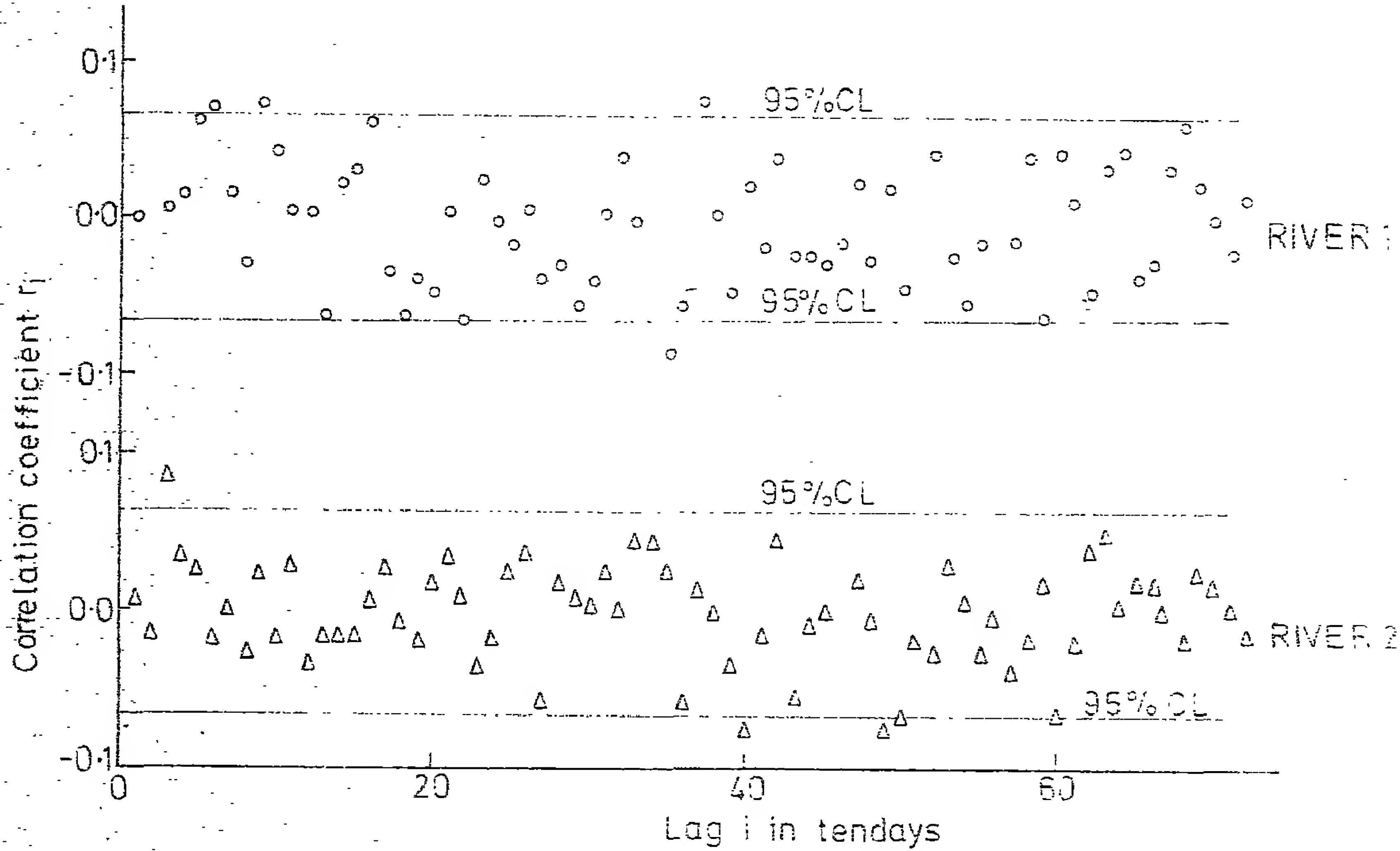


FIG. 5·3(a) CORRELOGRAM OF MULTIVARIATE TENDAILY RESIDUALS (MM) STATIONARY MODEL

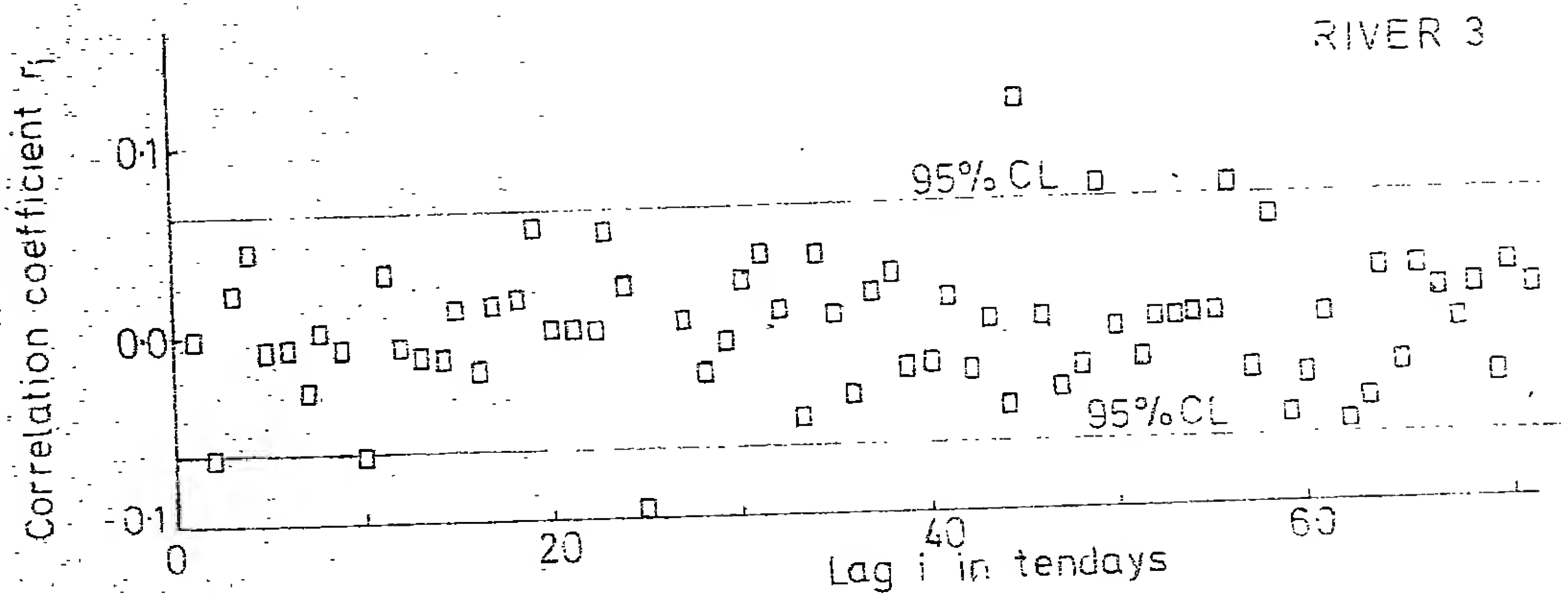


FIG. 5·3(a) CORRELOGRAM OF MULTIVARIATE TENDAILY RESIDUALS (MM) STATIONARY MODEL

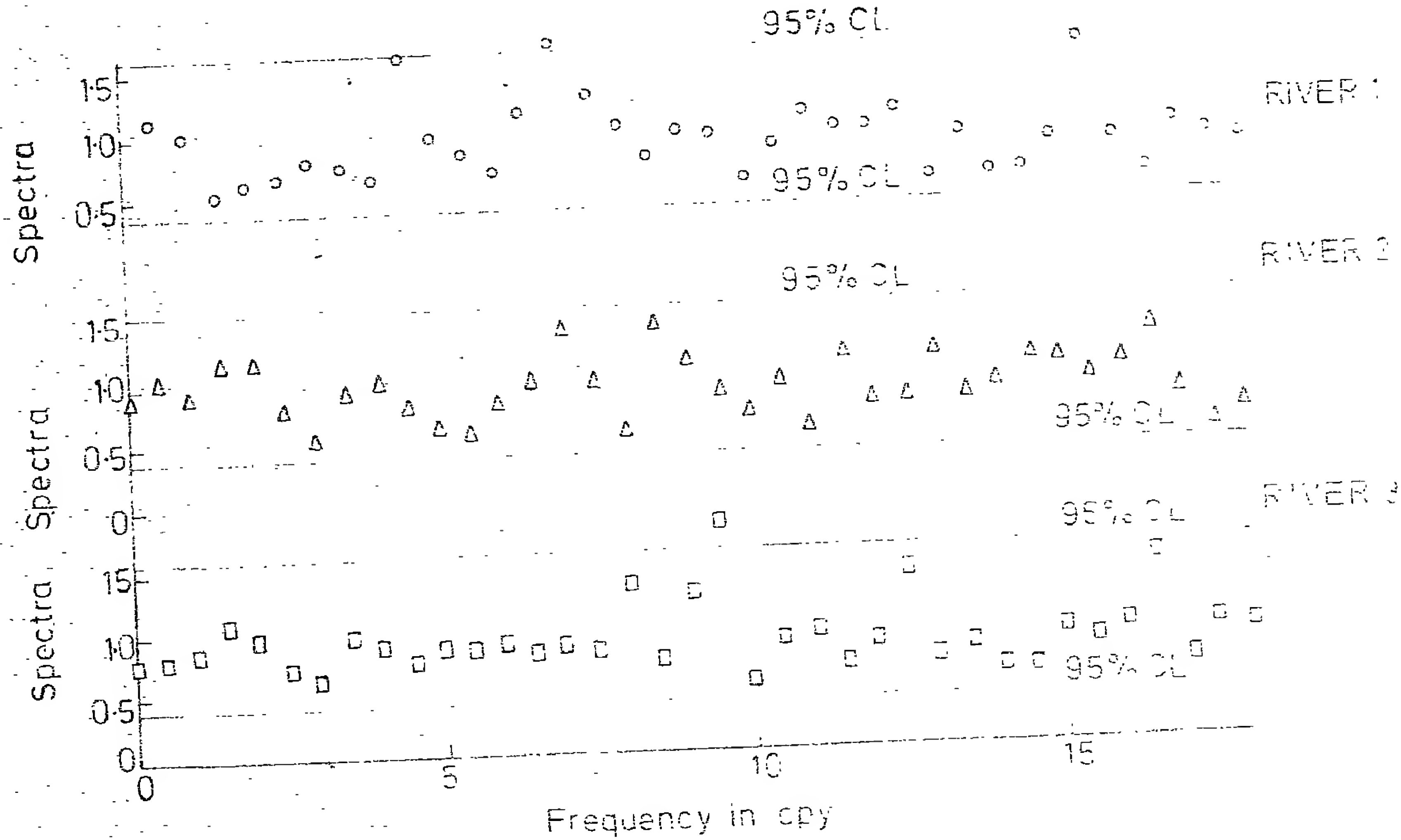


FIG. 5·3 (b) POWER SPECTRA OF MULTIVARIATE TENDAILY RESIDUALS MM (STATIONARY MODEL)

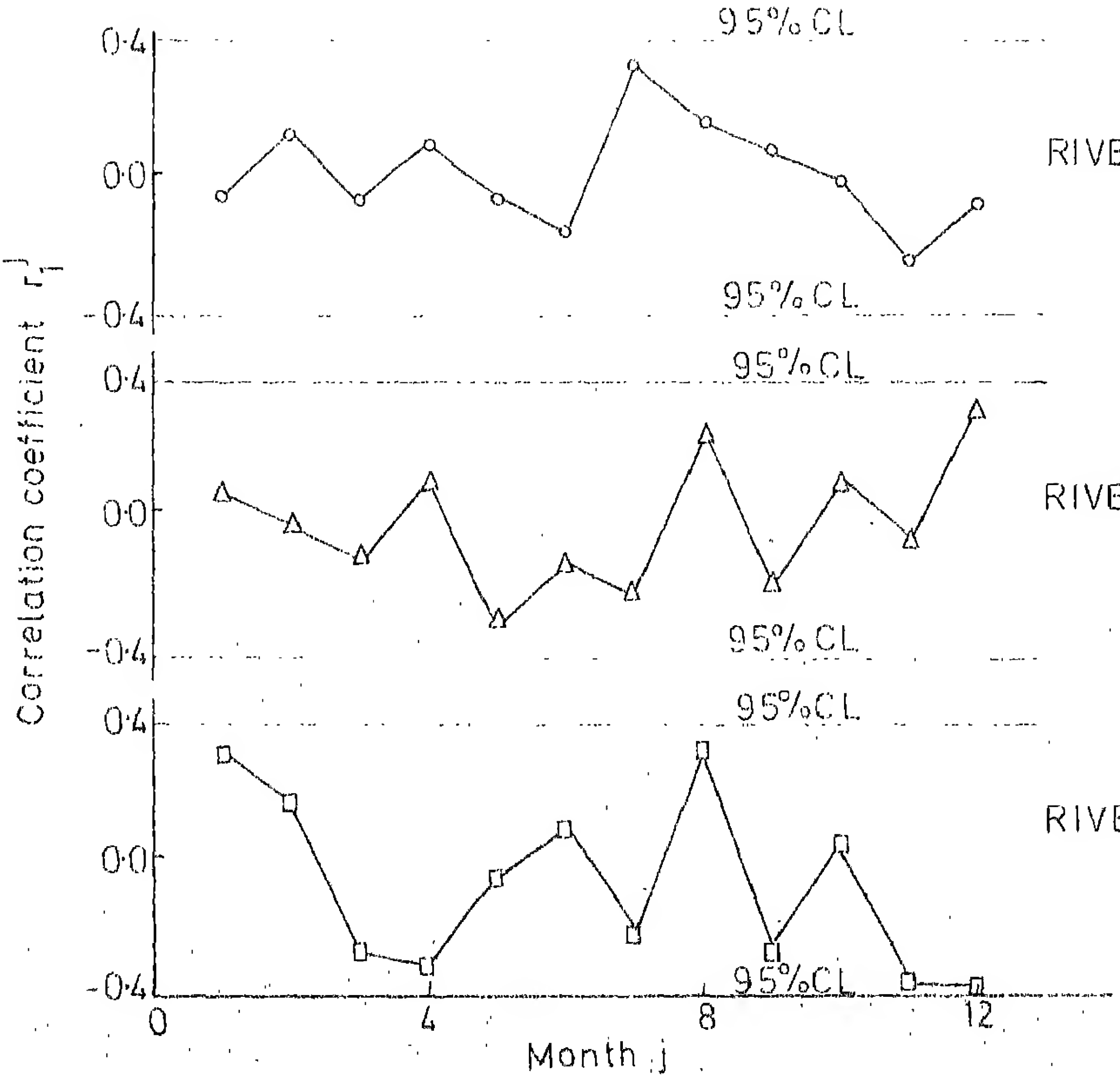


FIG. 5·4(a) SEASONAL SERIAL CORRELATION COEFFICIENTS OF MULTIVARIATE MONTHLY RESIDUALS (MLS) STATIONARY MODEL

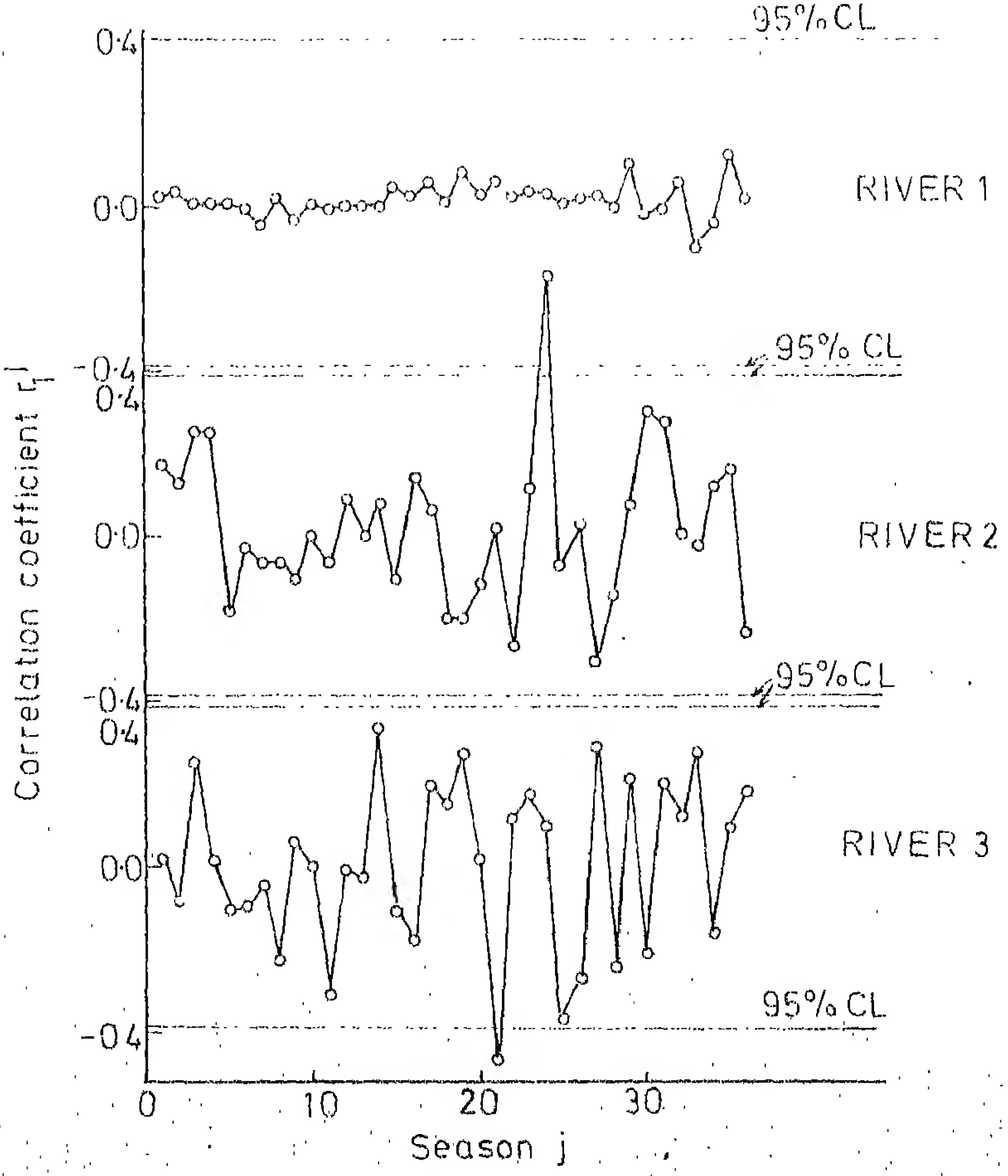


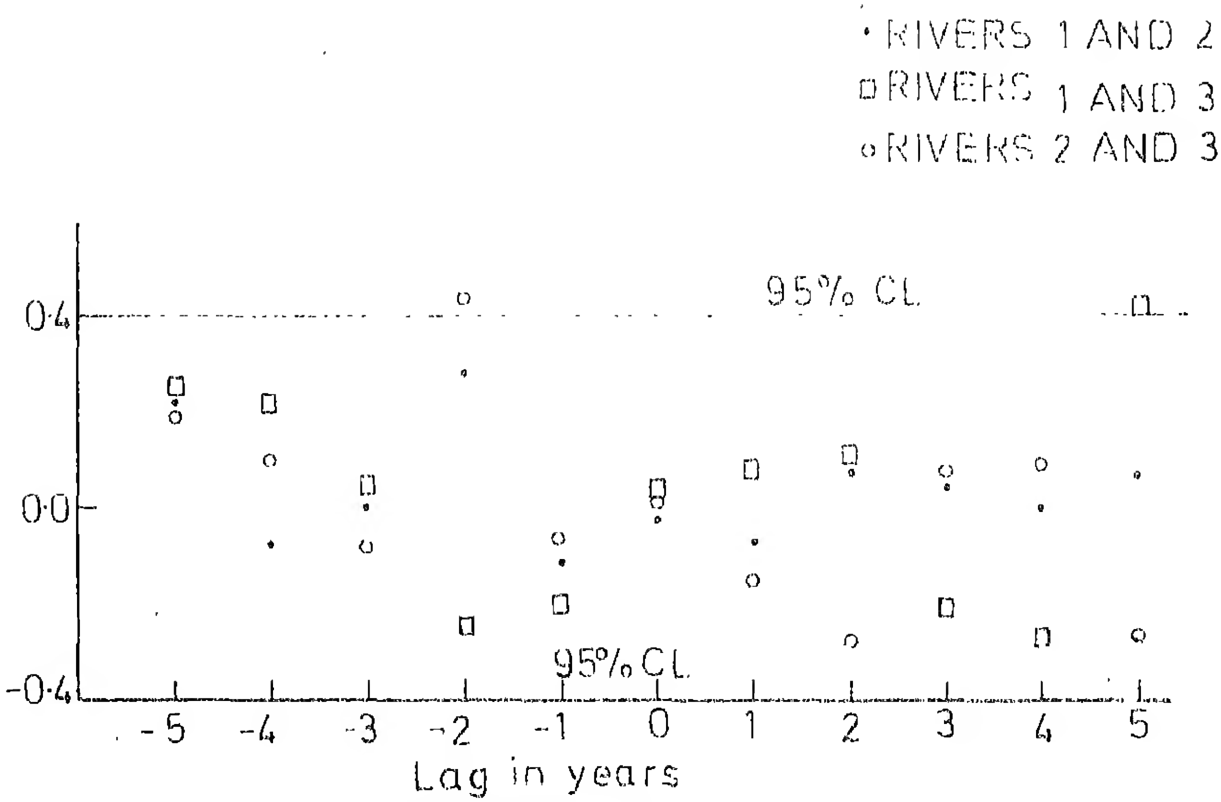
FIG.5·4(b) SEASONAL SERIAL CORRELATION COEFFICIENT OF MULTIVARIATE TENDAILY RESIDUALS (MLS) STATIONARY MODEL

the series taking two series at a time were estimated. They were tested for their significance at 95% level. Using the coherence spectra between any two time series, the independence among the time series was tested in the frequency domain. The significance of these spectra was ascertained by the procedure due to Granger and Hatanka (1964). If the sample size is N and the cross spectrum is estimated over m frequency bands, then the distribution of coherence when the true coherence is zero at the frequency f is given by

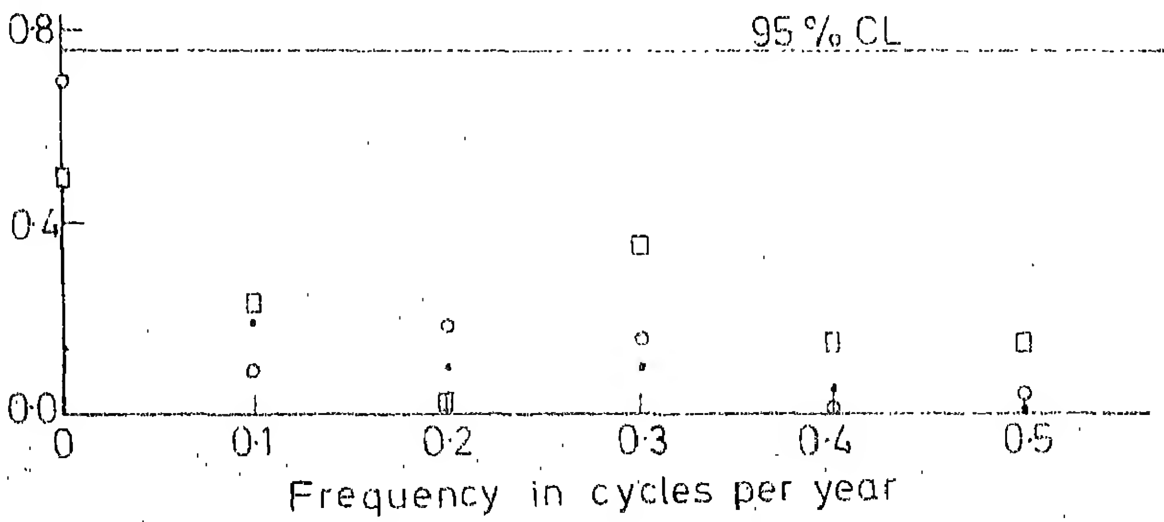
$$F(f) = 1 - (1 - f^2)^{N/m} \tag{5.14}$$

They have also provided a table giving the least significant values of the estimated coherence spectra for various values of N/m. This was used in the present study to determine if the estimated values were significant at 95% level. Furthermore, for the monthly and tendaily multivariate residual series $\{\eta(t)\}$, the lagzero crosscorrelation coefficient for each season was tested for significance at 95% confidence level.

The cross correlogram and the coherence spectra for the annual series taken in pairs are shown in Fig. 5.5(a) and Fig. 5.5(b) respectively, along with the 95% confidence limits. The crosscorrelation and the coherence between the time series are found to be not significantly



G·5·5(a) CROSSCORRELOGRAM OF MULTIVARIATE ANNUAL RESIDUALS



G·5·5(b) COHERENCE SPECTRA OF MULTIVARIATE ANNUAL RESIDUAL

different from zero. The cross correlogram and the coherence spectra for the monthly series are shown in Fig . 5.5(c) and Fig. 5.5(d) respectively along with their 95% confidence limits. They indicate that the $\{\eta(t)\}$ series may be considered to be a pure random time series. The seasonwise cross-correlation coefficients between any two monthly residual series are plotted in Fig. 5.6. Between stations 1 and 2 only MLE indicates a month with significant crosscorrelation. Between stations 1 and 3, all the three methods indicate one month with significant crosscorrelation among the residuals. Between rivers 2 and 3, MM and MLS indicate two months with significant crosscorrelation. The cross correlograms between the tendaily $\{\eta(t)\}$ series of stations 1 and 2, 1 and 3 and 2 and 3 shown in Fig. 5.7. The coherence spectra are shown in Fig. 5.8. They indicate that these series can be regarded as mutually independent at 95% CL. Seasonwise crosscorrelation coefficients for the three methods of standardisation viz., MM, MLE and MLS are shown in Figs. 5.9(a), (b) and (c). MM seems to yield residuals with the least number of significant seasonal crosscorrelation coefficients, closely followed by MLS. The number of seasons when these coefficients are significantly different from zero are 4,7 and 4 between stations 1 and 2; 4, 4 and 4 between stations 1 and 3; and 7, 10 and 9 between stations 2 and 3, for MM, MLE and MLS respectively. Hence the $\{\eta(t)\}$ series retain significant crosscorrelation in atleast 11 to 20 of the seasons.

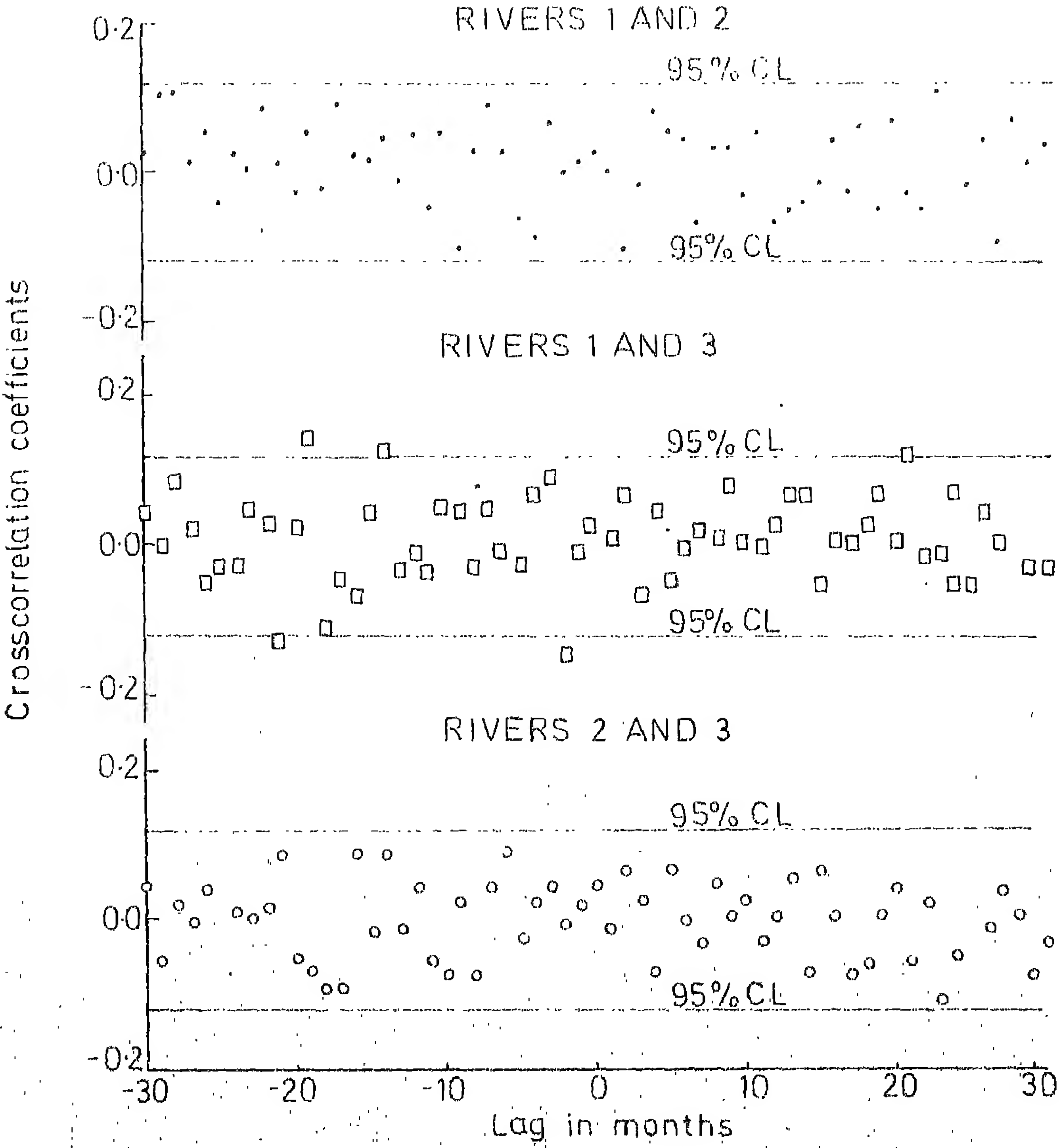


FIG. 5·5(c) CROSS CORRELOGRAM OF MULTIVARIATE MONTHLY RESIDUALS (MM) STATIONARY MODEL

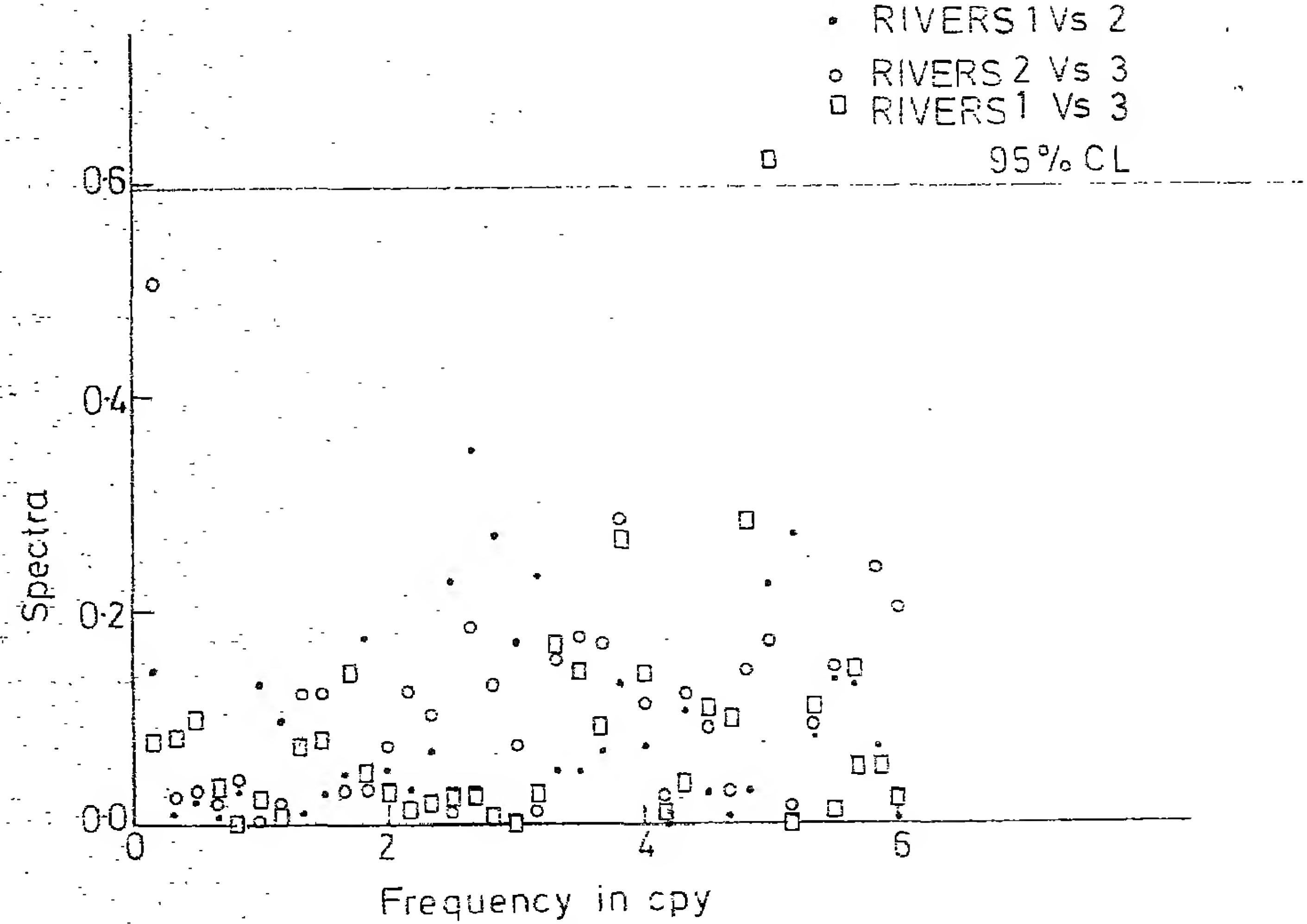


FIG. 5·5 (d) COHERENCE SPECTRA OF MULTIVARIATE MONTHLY RESIDUALS (MM) STATIONARY MODEL

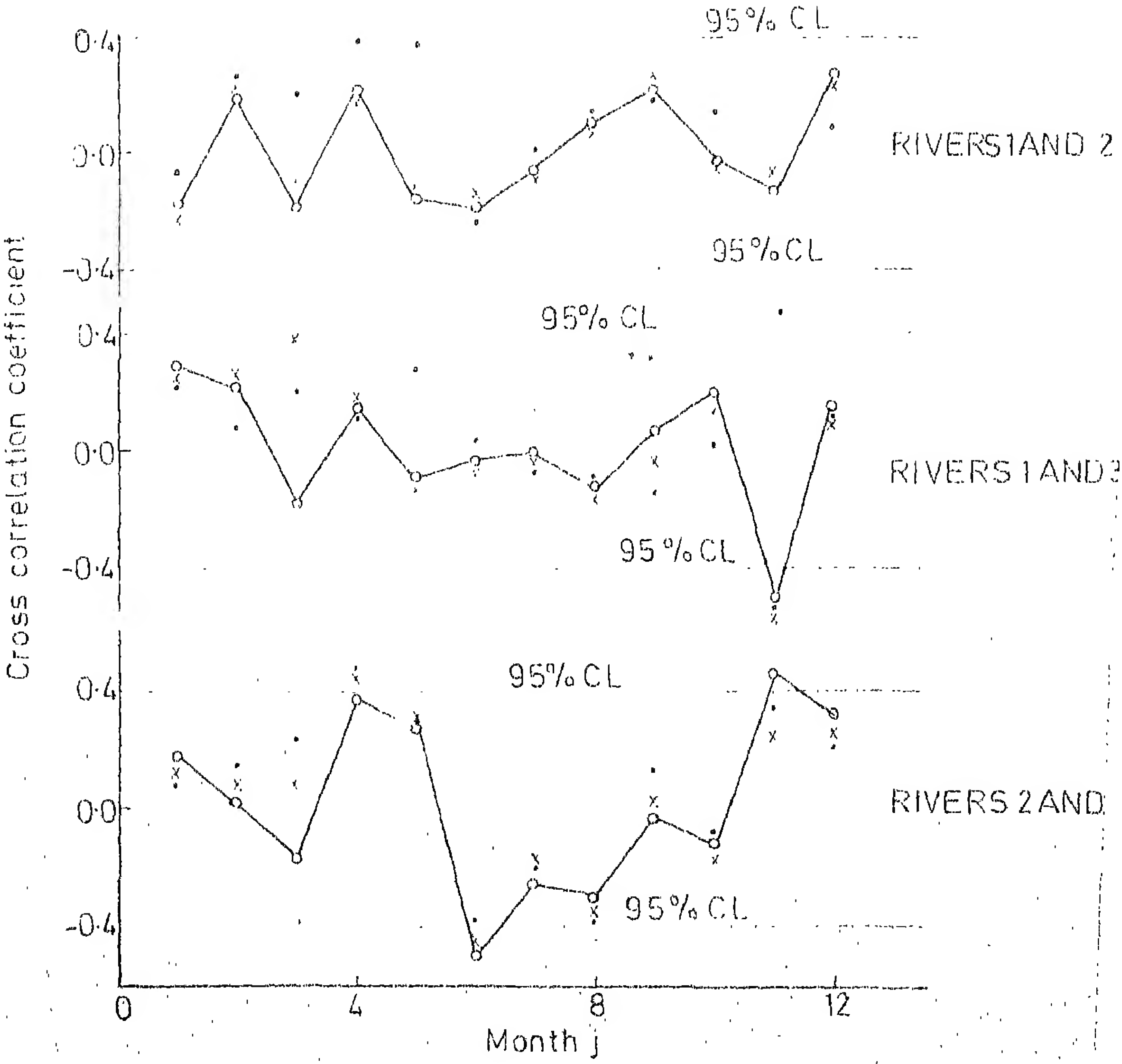


FIG. 5.6 SEASONAL CROSSCORRELATION COEFFICIEN OF MULTIVARIATE MONTHLY RESIDUALS STATIONARY MODEL

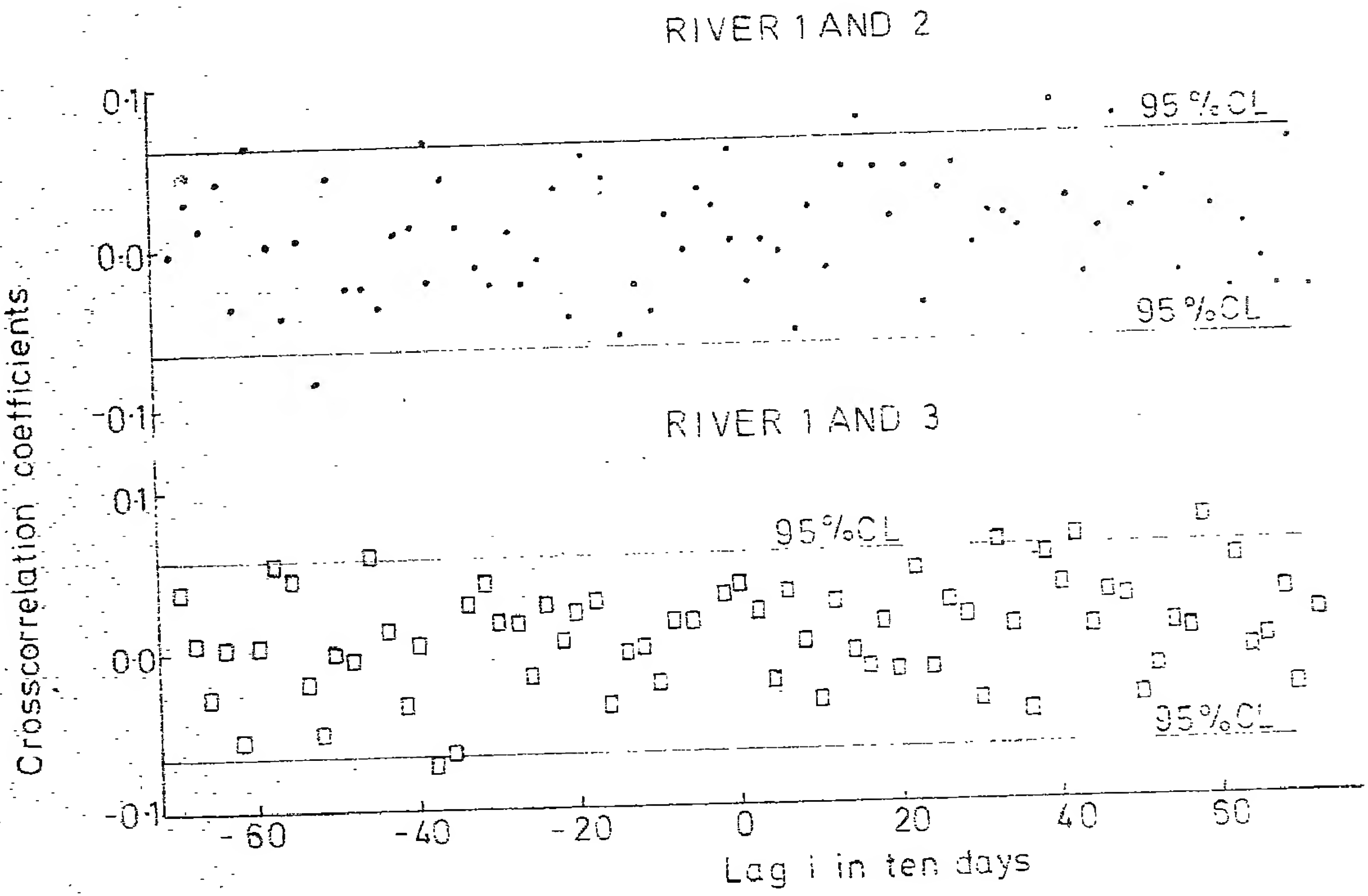
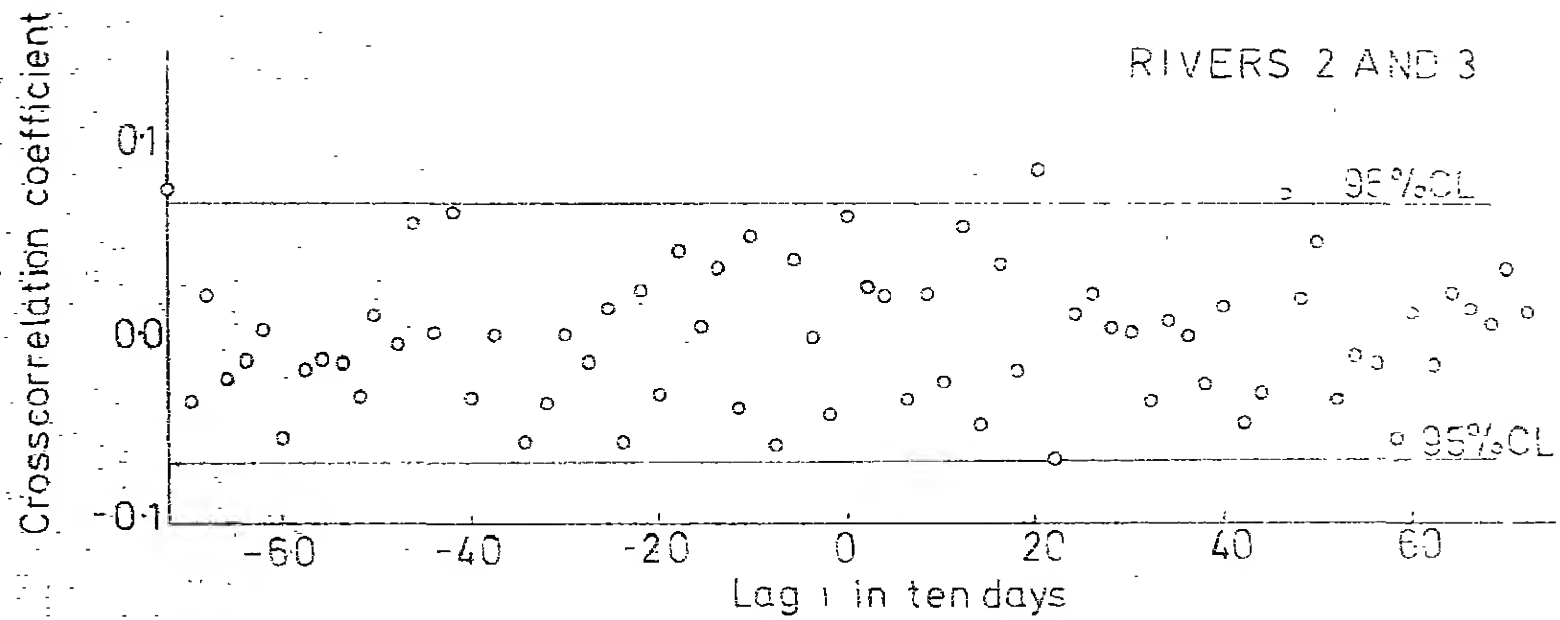


FIG. 5·7 CROSS CORRELOGRAM OF MULTIVARIATE TENDAILY RESIDUALS (MM) STATIONARY MODEL



FIG·5·7 CROSS CORRELOGRAM OF MULTIVARIATE TENDAILY RESIDUALS (MM) STATIONARY MODEL

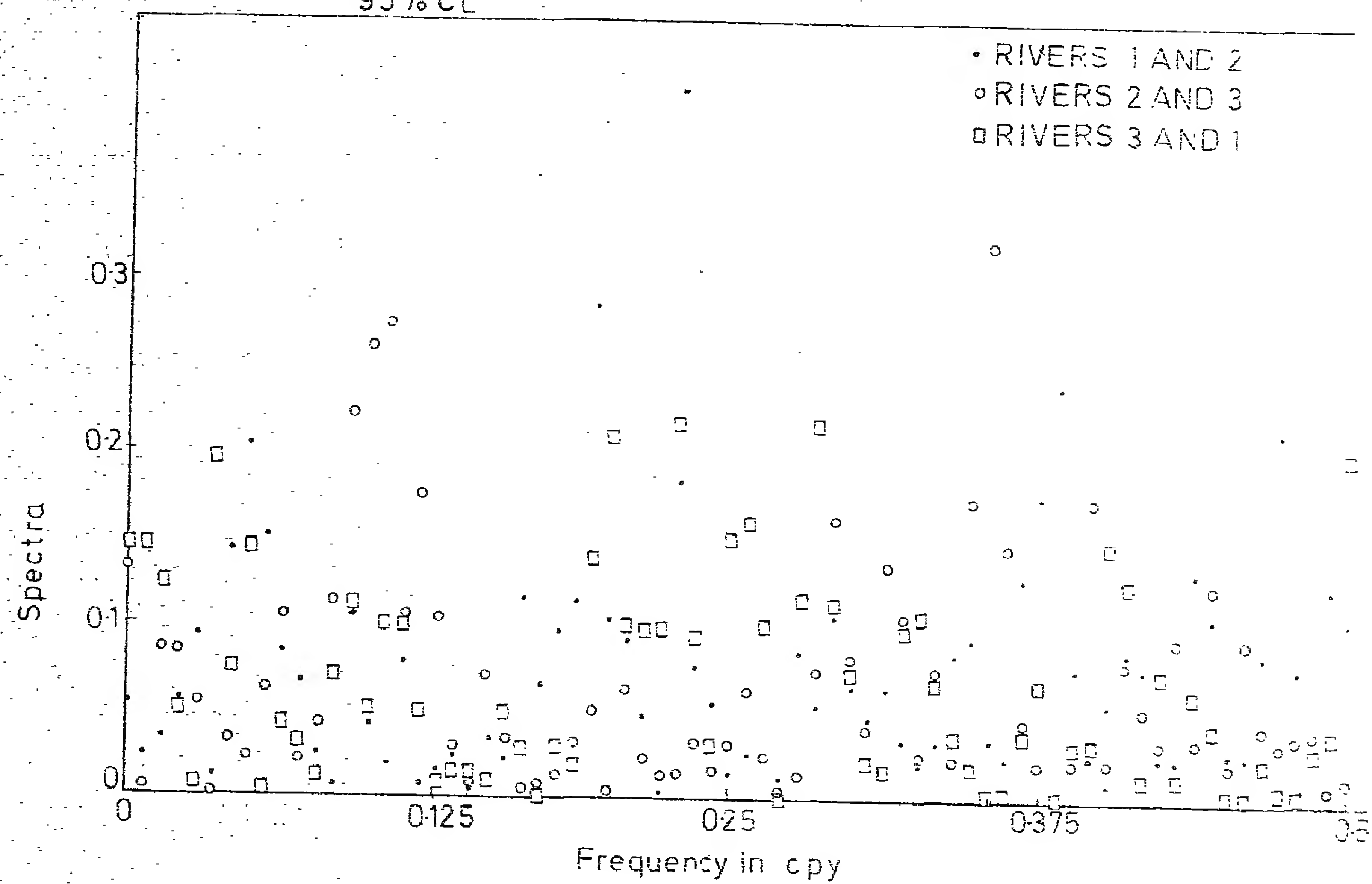
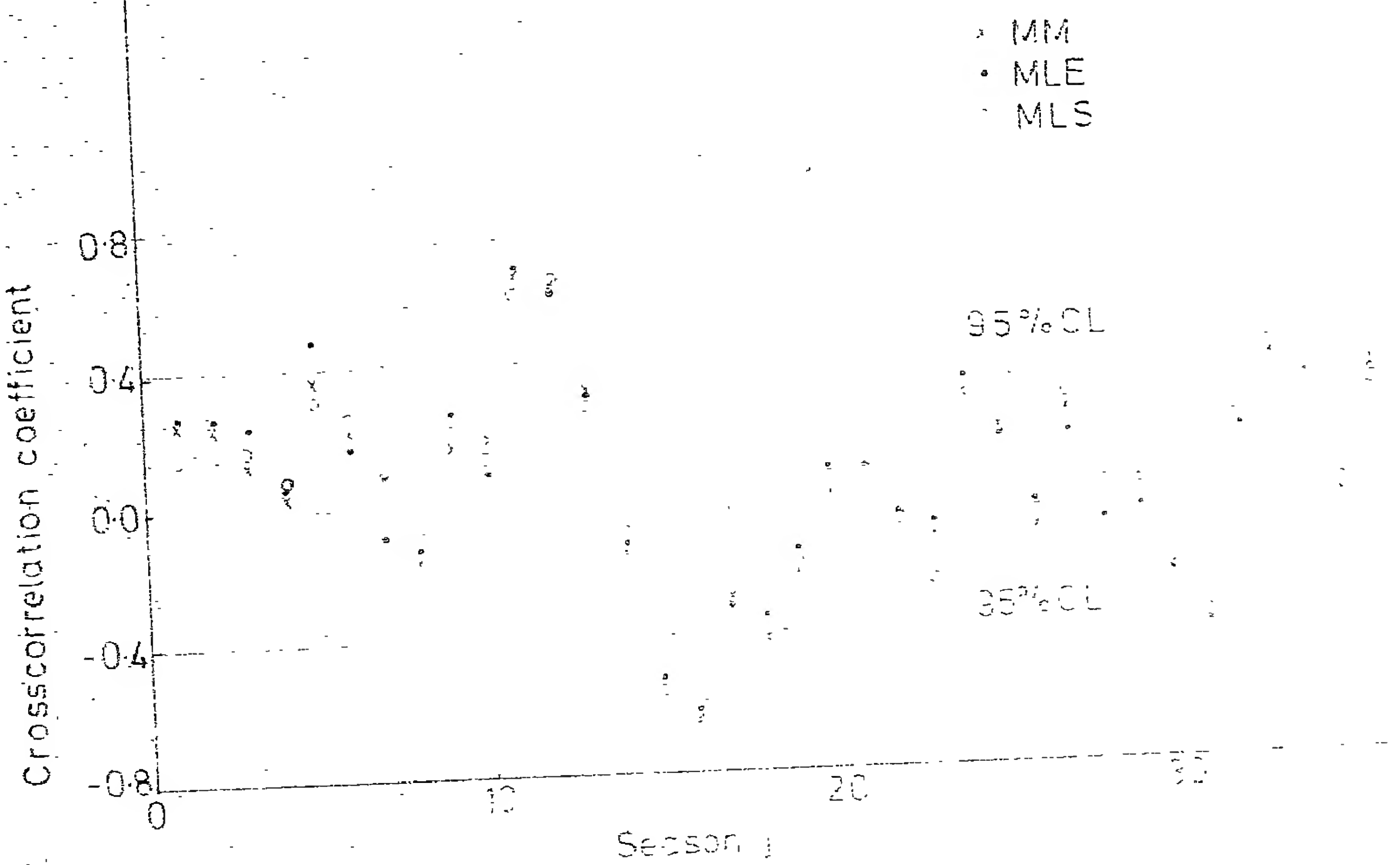


FIG. 5·8 COHERENCE SPECTRA OF MULTIVARIATE TENDAILY RESIDUALS (MM) STATIONARY MODEL



FIG·5·9(a) SEASONAL CROSSCORRELATION COEFFICIENTS OF MULTIVARIATE TENDAILY RESIDUALS (STATIONARY MODEL) RIVERS 1 AND 2

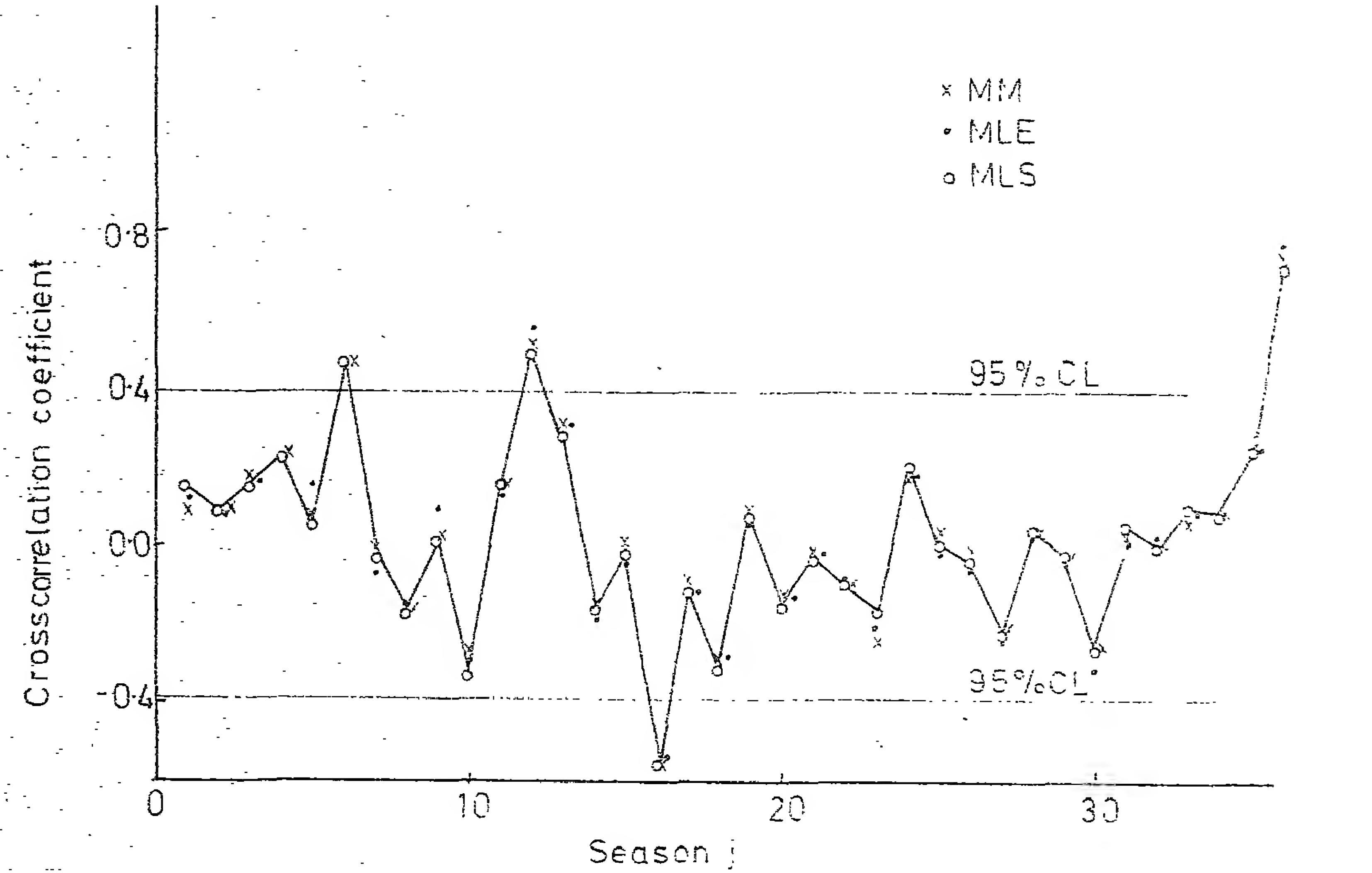


FIG. 5·9 (b) SEASONAL CROSSCORRELATION COEFFICIENTS OF TENDAILY RESIDUALS (STATIONARY MODEL) RIVERS 1 AND 3

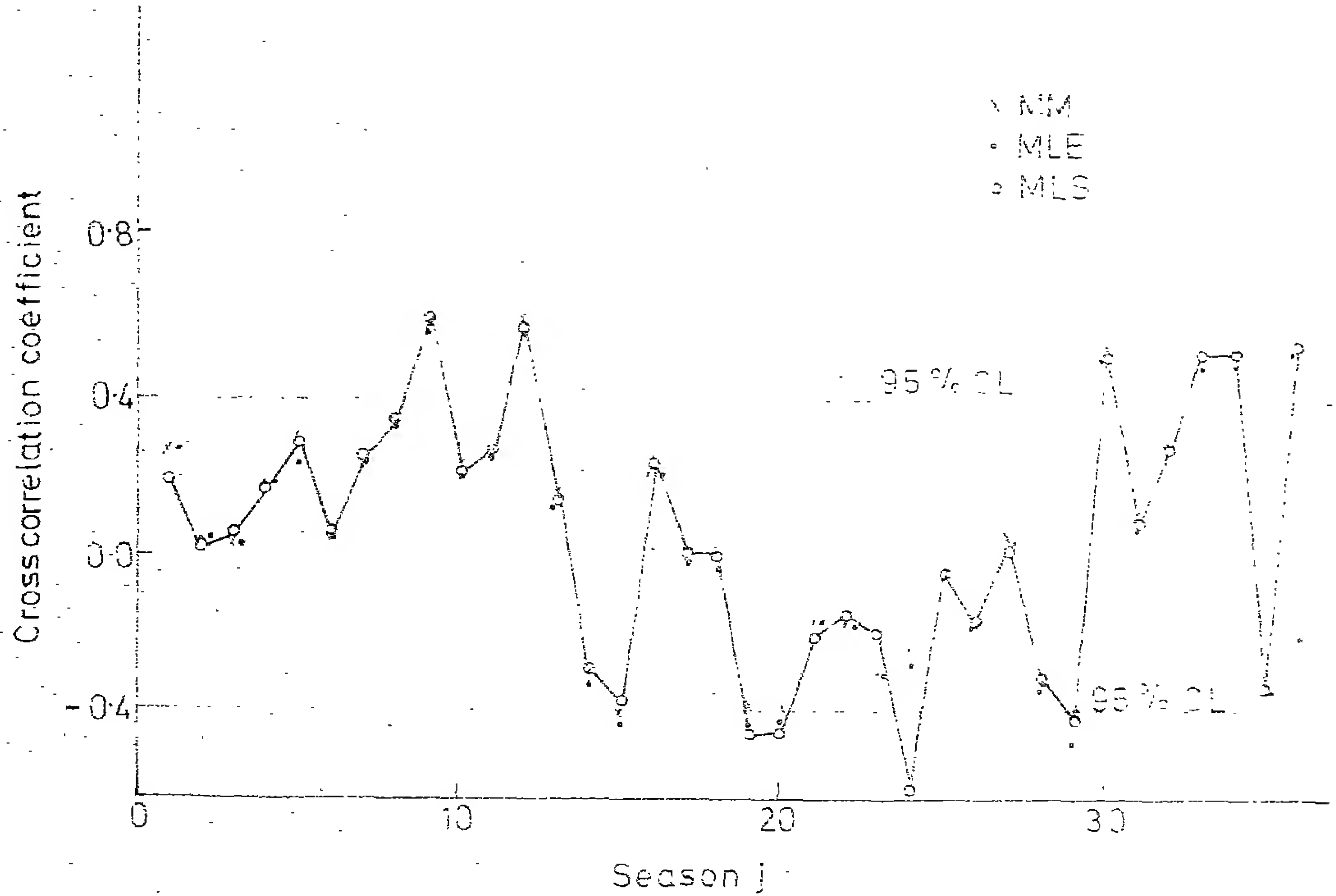


FIG. 5·9 (c) SEASONAL CROSS CORRELATION COEFFICIENTS OF TENDAILY RESIDUALS STATIONARY MODEL RIVERS 2 AND 3

TABLE 5.6 χ^2 ESTIMATES OF MULTIVARIATE TENDAILY RESIDUALS (STATIONARY ZEROETH ORDER MODEL)

NUMBER OF SEASONS WHEN HYPOTHESIS IS REJECTED AT 95% LEVEL

RIVER	STANDARDISATION BY		
	MM	MLE	MLS
1	9	9	7
2	1	1	2
3	3	3	4

5.2.4 Discussion of results for stationary multivariate models

The results of fitting a stationary multivariate model for the annual, monthly and tendaily series indicate the following:

i) The stationary multivariate first order AR model represents the annual series adequately and the residuals constitute independent random series, and there is no significant coherence between the residuals.

ii) The monthly and tendaily series can be represented by a multivariate zeroeth order AR model with serially independent residuals. This agrees with the conclusion of Yevjevich (1975) for the net basin supply to the Great Lakes of North America. But there is a significant cross-correlation between the residuals for one or two months

out of 12 in the case of monthly series, and a significant number of seasons for the tendaily series.

iii) In case the significant seasonal crosscorrelation between the residual series is ignored, the stationary multivariate AR model can be considered to be a satisfactory representation of the univariate residuals in terms of independent random components.

5.3 Decoupled Nonstationary Multivariate Models

A stationary multivariate model is found to be satisfactory for the annual time series. Stationary multivariate models for monthly and tendaily series indicate significant seasonwise crosscorrelation between the multivariate residual series. It hence seems necessary to consider nonstationary multivariate models for the process. It is proposed to use a first order AR model with seasonally variable [C] and [D] matrices.

5.3.1 Estimation of parameters

The standardised univariate residuals $E(t)$ of the monthly and tendaily series, corresponding to MM, MLE and MLS were used in the multivariate modelling of the time series. The coefficient matrices $[C^j]$ and $[D^j]$ where j is the seasonal index were estimated using the procedure in Sec.5.1.

It may be noted that the sample size for estimation of parameters of the nonstationary model is N/s as against N for the stationary model. The estimates of the coefficient matrices for monthly residuals are given in Table 5.7. For tendaily series, it was not possible to fit a nonstationary multivariate AR model for 4 seasons out of 36, viz., during seasons with serial numbers 11, 24, 30 and 36. This was because, during these seasons, $[D^j]\ [D^j]^T$ was not positive definite implying that the variance explained by the $[C^j]$ matrix was larger than the original variance and hence the $[D^j]$ had to explain a negative variance. This is not possible as the elements of $[D^j]$ are assumed to be real quantities. Matalas and Wallis (1971) have indicated that under this conditions, the Young and Pisano model cannot be used to represent the data. It may be possible to use other models or else to use procedures suggested by Fiering (1968), Crosby and Maddock (1970) etc., in the case of inconsistent matrices. This was considered to be outside the scope of this study. Hence whenever such an inconsistency was met with, the simpler zeroeth order multivariate model (Eq.5.12) which corresponds to that of Yevjevich (1975) was used. This will be valid as $[M_o^j]$ is positive definite.

The $[C^j]$ matrices obtained from the nonstationary model can be considered as realisations of a stationary [C] matrix. It is desirable to test whether the $[C^j]$ matrices

TABLE 5.7 PARAMETER ESTIMATES FOR NONSTATIONARY FIRST ORDER MULTIVARIATE MODEL (MONTHLY SERIES)

$F_{THEORETICAL,\ 5\%} = 9.3$ $Z_{THEORETICAL,} = 7.82$

MONTH	STANDARDISATION BY								
	MM			MLE			MLS		
1	-.13	-.45	.31	.01	-.21	.05	-.07	-.38	-.20
	.05	-.05	-.03	-.24	.00	.12	.07	-.01	-.09
	-.09	-.53	-.39	-.37	-.46	.53	-.02	-.38	.19
	Z_E = 1.7			1.5			1.8		
	F_E = 0.4			0.4			0.5		
2	.13	-.22	.18	.01	-.22	.20	.20	-.13	-.01
	.35	-.20	-.09	-.04	-.01	.02	-.31	-.11	-.21
	-.14	-.05	.21	.28	.03	.19	-.01	-.04	.12
	Z_E = 3.3			4.2			1.0		
	F_E = 0.8			1.1			0.3		
3	-.16	.39	-.23	-.35	.25	-.06	-.06	.28	-.16
	-.41	-.07	.14	-.54	.20	.04	-.33	-.04	-.21
	-.12	.42	-.38	-.28	.58	-.40	-.02	-.16	-.11
	Z_E = 8.4+			8.8+			3.7		
	F_E = 2.1			2.2			0.9		

contd....

MONTH	STANDARDISATION BY MM			MLE			MLS		
4	-.05	.44	-.49	-.45	.96	-.57	.02	-.03	-.01
	-.15	1.03	-.88	-.31	1.19	-.90	-.09	.42	-.25
	.13	1.00	-.98	.21	.87	-.93	-.04	.55	-.44
	Z_E =	9.8^+			10.4^+			1.8	
	F_E =	2.5			2.6			0.5	
5	-.26	.13	.25	-.63	.08	.55	-.27	.13	.28
	.11	-.36	.33	-.41	-.08	.46	.11	-.39	.42
	.31	-.41	.26	-.20	-.12	.27	.28	-.16	.09
	Z_E =	5.5			6.2			3.7	
	F_E =	1.8			1.6			0.9	
6	-.12	-.23	-.05	-.12	-.15	-.14	-.16	-.22	-.04
	.20	-.35	-.48	-.12	.23	-.37	.15	-.36	-.39
	.35	-.41	-.28	-.07	-.24	-.15	.32	-.36	-.27
	Z_E =	7.2			8.3^+			7.0	
	F_E =	1.8			2.1			1.8	
7	.60	-.26	-.20	.69	-.33	-.29	.61	-.27	-.23
	.86	-.40	-.16	.98	-.49	-.18	.88	-.40	-.24
	.52	-1.0	-.38	.62	-.17	-.40	.51	-.20	-.33
	Z_E =	9.0^+			9.2^+			8.6^+	
	F_E =	2.3			2.3			2.2	

Contd...

MONTH	STANDARDISATION BY								
	MM			MLE			MLS		
8	.24	-.03	-.07	.22	.01	-.08	.25	-.03	-.10
	-.40	.45	.25	-.51	.54	.31	-.37	-.44	.21
	.16	-.36	.25	.23	-.42	.21	.20	-.45	.24
	Z_E = 8.5^+			10.4^+			8.0^+		
	F_E = 2.1			2.6			2.0		
9	.36	-.88	.61	.24	-.68	.51	.28	-.70	.49
	.67	-.72	.56	.54	-.60	.55	.61	-.53	.38
	.05	-.03	-.07	.08	-.04	-.07	-.01	-.07	-.02
	Z_E = 14.5^+			16.2^+			13.7^+		
	F_E = 3.6			4.1			3.9		
10	.32	-.11	-.66	.44	-.31	-.62	.30	-.13	-.60
	.25	.03	-.32	.46	-.10	-.37	.23	.00	-.27
	.49	-.08	-.30	.66	-.22	-.33	.49	-.09	-.30
	Z_E = 6.0			6.0			5.5		
	F_E = 1.5			1.5			1.4		
11	.05	-.63	.46	.16	-.03	.19	.17	-.81	.63
	.49	-.49	.52	.18	-.17	.33	.60	-.67	.68
	.20	.26	.00	.14	.37	-.05	.21	.19	.02
	Z_E = 10.1^+			11.5^+			9.0^+		
	F_E = 2.5			2.9			2.3		

Contd...

MONTH	STANDARDISATION BY								
	MM			MLE			MLS		
12	.33	-.24	-.31	-.48	.66	-.34	.28	-.31	-.18
	-.25	.52	.23	-.39	.56	-.17	-.28	.52	-.22
	-.19	-.05	.33	-.46	.07	.42	-.08	.26	.04
	Z_E = 11.8$^+$			12.7$^+$			11.0$^+$		
	F_E = 3.0			3.2			2.8		

+ indicates significant values

[D] MATRIX

MONTH	STANDARDISATION BY									IS $[DD^T]$ PD ?
	MM			MLE			MLS			
1	.94			.98			.95			
	.39	.92		.56	.81		.39	.92		YES
	.41	.58	.59	.61	.37	.53	.40	.54	.68	
2	.99			.99			.99			
	.73	.64		.76	.64		.73	.65		YES
	.66	.38	.63	.65	.32	.66	.61	.26	.74	
3	.96			.96			.96			
	.50	.78		.61	.68		.48	.77		YES
	.56	.53	.55	.58	.43	.57	.15	.01	.97	

Contd...

MONTH	STANDARDISATION BY									IS $[DD^T]$ PD?
	MM			MLE			MLS			
4	.97			.94			1.0			
	.64	.61		.75	.47		.73	.64		YES
	.45	.52	.44	.57	.42	.41	.51	.53	.65	
5	.93			.92			.92			
	.48	.84		.74	.59		.45	.83		YES
	.29	.60	.68	.75	.16	.61	.27	.57	.74	
6	.95			.96			.95			
	.33	.78		.24	.82		.37	.80		YES
	.25	-.35	.79	.39	-.29	.81	.24	-.36	.81	
7	.88			.84			.87			
	.31	.67		.26	.60		.29	.66		YES
	.25	.14	.83	.19	.06	.82	.22	.01	.87	
8	.98			.98			.98			
	.73	.57		.77	.50		.73	.58		YES
	.40	-.10	.84	.48	-.24	.76	.38	-.07	.83	
9	.79			.86			.83			
	.50	.55		.50	.60		.58	.56		YES
	.59	.56	.58	.52	.56	.64	.56	.35	.74	

Contd...

MONTH	STANDARDISATION BY									IS [DDT] PD?
	MM			MLE			MLS			
10	.80			.78			.83			
	.50	.82		.57	.72		.52	.82		YES
	.49	.20	.75	.45	.09	.74	.48	.20	.75	
11	.85			.98			.82			
	.40	.81		-.07	.96		.36	.80		YES
	-.05	.60	.70	-.38	.49	.68	-.09	.71	.62	
12	.93			.89			.95			
	.84	.46		.60	.71		.84	.46		YES
	.74	.28	.56	.65	.41	.51	.63	.31	.69	

are significantly different from the stationary [C] matrix derived for a stationary model. Let Z^j be the Z statistic estimated for the j-th season (Subsec. 5.2.2). Z^j and Z are approximately distributed according to chisquare distribution and with the same number of degrees of freedom K. Hence the statistic F given by Z^j/Z will be approximately distributed according to F distribution with (K, K) degrees of freedom. If the estimate of F exceeds the theoretical value at, say, 95% confidence level, then the hypothesis that F is significantly different from unity is accepted. This is considered to imply that the $[C^j]$ matrix is statistically different from the [C] matrix at the same level.

5.3.2 Significance tests on parameters

Validation of the nonstationary model was done using the procedure of Subsec. 5.2.2. The $[D^j][D^j]^T$ matrix was tested for each j for positive definiteness. Z^j statistic was also estimated for all j and used to test whether the $[C^j]$ matrices were significant for each j. The F statistic was used to test whether the $[C^j]$ matrices of the nonstationary first order model were significantly different from the [C] matrix of the stationary first order model. For a zeroeth order stationary model the test on F statistic is inapplicable.

Monthly series: $[D^j][D^j]^T$ matrices were found to be positive definite for all months of the year. The results of the tests on $[C^j]$ matrices are given in Table 5.7. They indicate that in about 7 of the 12 months, $[C^j]$ is significantly different from zero. Results of the F test show that the above $[C^j]$ matrices are not significantly different from the [C] matrix of the stationary first order model (Sec. 5.2).

Tendaily series: The results of the tests on tendaily series are given in Table 5.8. $[D^j][D^j]^T$ matrices were not positive definite for 4 out of 36 seasons. As indicated in Subsec. 5.3.1, during these seasons, the simpler zeroeth order multivariate model was used. The test for Z statistic is not applicable for these seasons. Results of the Z test showed that, out of the remaining 32 seasons, $[C^j]$ was found to be significantly different from zero during 11 seasons. Results of the F test showed that the $[C^j]$ matrices for the nonstationary first order model were not significantly different from the [C] matrix of the stationary first order model.

5.3.3 Validation of the fitted nonstationary models

Normality of residuals: The results of the chisquare test on the residuals for the monthly series are given in

TABLE 5.8 Z AND F STATISTICS FOR NONSTATIONARY MULTIVARIATE MODEL (TENDAILY SERIES)

SEASON	Z_{EST}	F_{EST}	ORDER OF MODEL$^{\alpha}$	SEASON	Z_{EST}	F_{EST}	ORDER OF MODEL$^{\alpha}$
1	1.3	0.2	1	19	4.4	0.8	1
2	3.7	0.6	1	20	3.7	0.6	1
3	7.3	1.2	1	21	21.2^{+}	3.5	1
4	2.6	0.4	1	22	2.1	0.4	1
5	4.3	0.7	1	23	10.8^{+}	1.8	1
6	12.1^{+}	2.0	1	24	0.0	0.0	0
7	0.9	0.2	1	25	18.7^{+}	3.1	1
8	8.6^{+}	1.4	1	26	2.4	0.4	1
9	3.0	0.5	1	27	19.6^{+}	3.3	1
10	1.4	0.2	1	28	13.2^{+}	2.2	1
11	0.0	0.0	0	29	4.2	0.7	1
12	26.8^{+}	4.8	1	30	0.0	0.0	0
13	0.4	0.1	1	31	5.1	0.9	1
14	2.7	0.6	1	32	49.0^{+}	8.2	1
15	16.6^{+}	2.8	1	33	19.5^{+}	3.3	1
16	5.4	0.9	1	34	1.7	0.3	1
17	4.0	0.7	1	35	4.5	0.8	1
18	6.2	1.0	1	36	0.0	0.0	0

\+ indicates significance at 95% level

α First order model used if $[DD^T]$ is positive definite otherwise zeroeth order model used.

Table 5.9. They may be compared with the corresponding results for a stationary model (Table 5.5) and the univariate residuals (Table 4.7). The results of the chisquare test on the residuals of the nonstationary multivariate model for the tendaily series are given in Table 5.10.

TABLE 5.9 χ^2 ESTIMATES OF MULTIVARIATE MONTHLY RESIDUALS (NONSTATIONARY FIRST ORDER MODEL)

NUMBER OF SEASONS WHEN HYPOTHESIS IS REJECTED AT 95% LEVEL

RIVER	STANDARDISATION BY		
	MM	MLE	MLS
1	2	2	2
2	0	0	0
3	0	0	0

TABLE 5.10 χ^2 ESTIMATES OF MULTIVARIATE TENDAILY RESIDUALS (NONSTATIONARY MODEL)

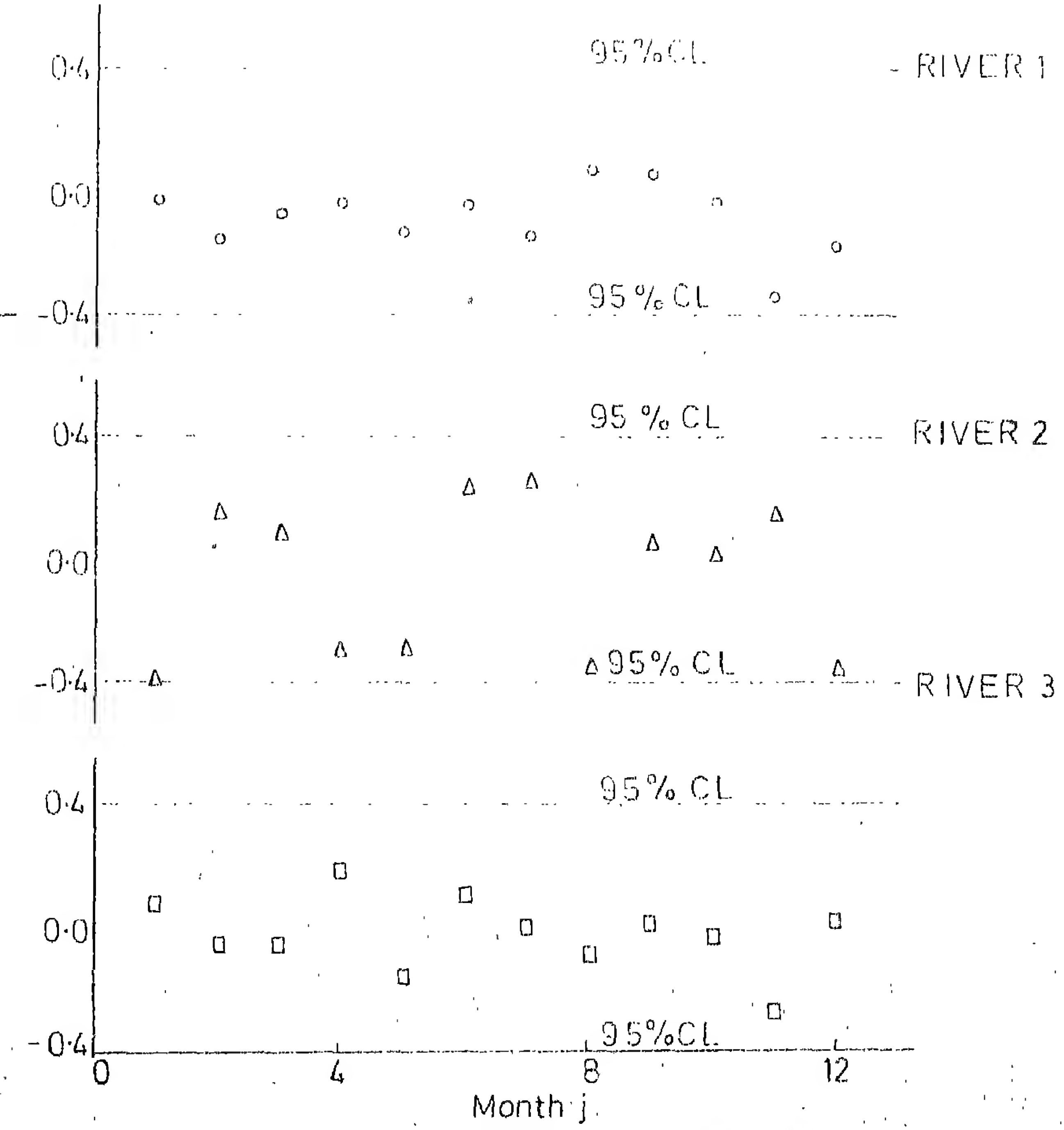
NUMBER OF SEASONS WHEN HYPOTHESIS IS REJECTED AT 95% LEVEL

RIVER	STANDARDISATION BY		
	MM	MLE	MLS
1	6	6	5
2	1	1	1
3	2	2	3

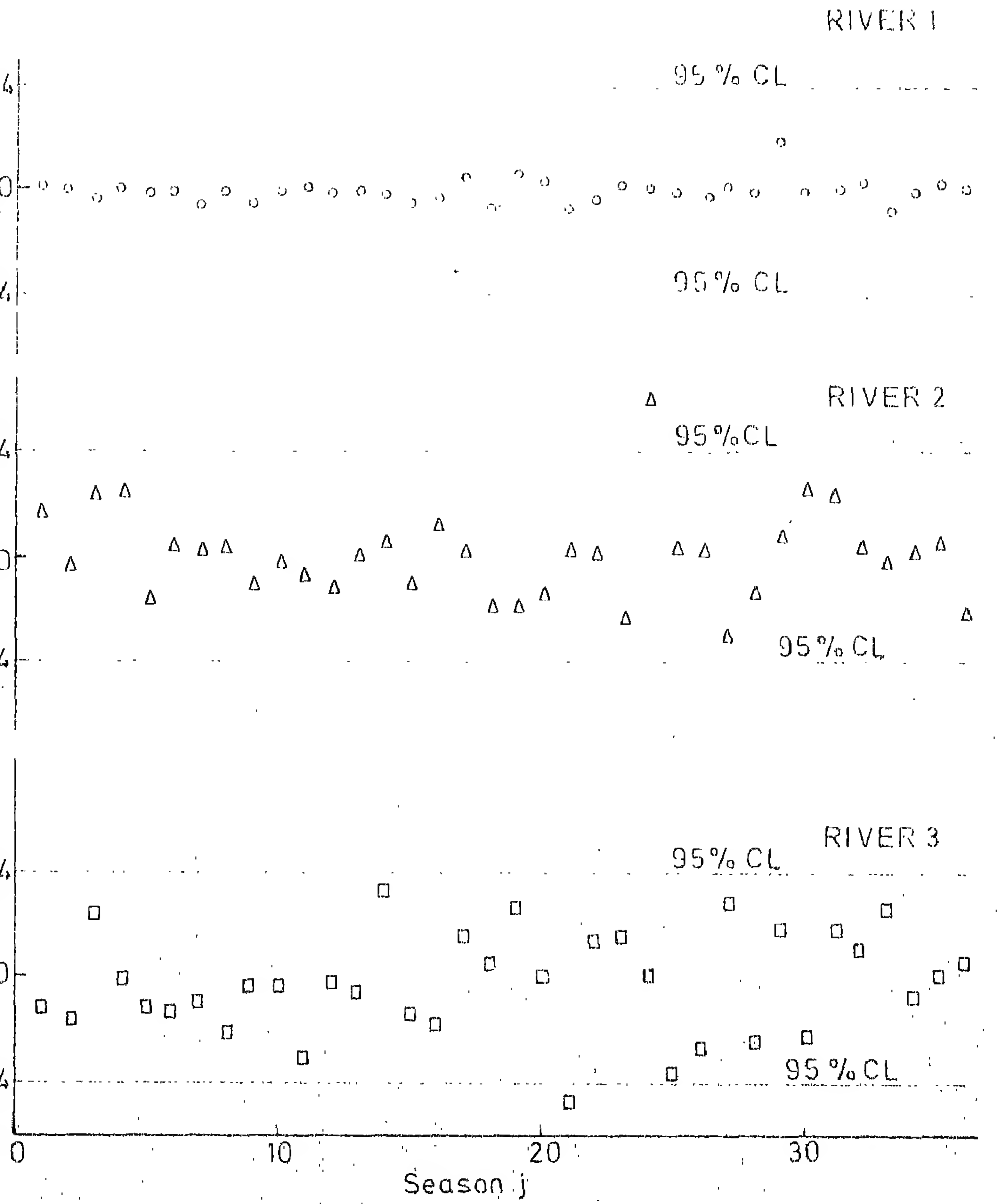
They may be compared with the corresponding results for the residuals of the stationary model (Table 5.6) and the univariate residuals (Table 4.7). The results indicate that for both monthly and tendaily data, the residuals from a nonstationary multivariate model are more nearly normally distributed than the residuals of the stationary multivariate model or the univariate residuals.

Independence within the time series: The correlogram and spectral analyses indicate that each of the multivariate residual series may be considered as pure random. Furthermore, the seasonal serial correlation coefficients are estimated for each series for each season and are shown in Fig. 5.10 for the monthly series and Fig. 5.11 for the tendaily series. In either case, they are not statistically significant at 95% level. Hence the residual series from the nonstationary multivariate model are not serially and seasonally correlated.

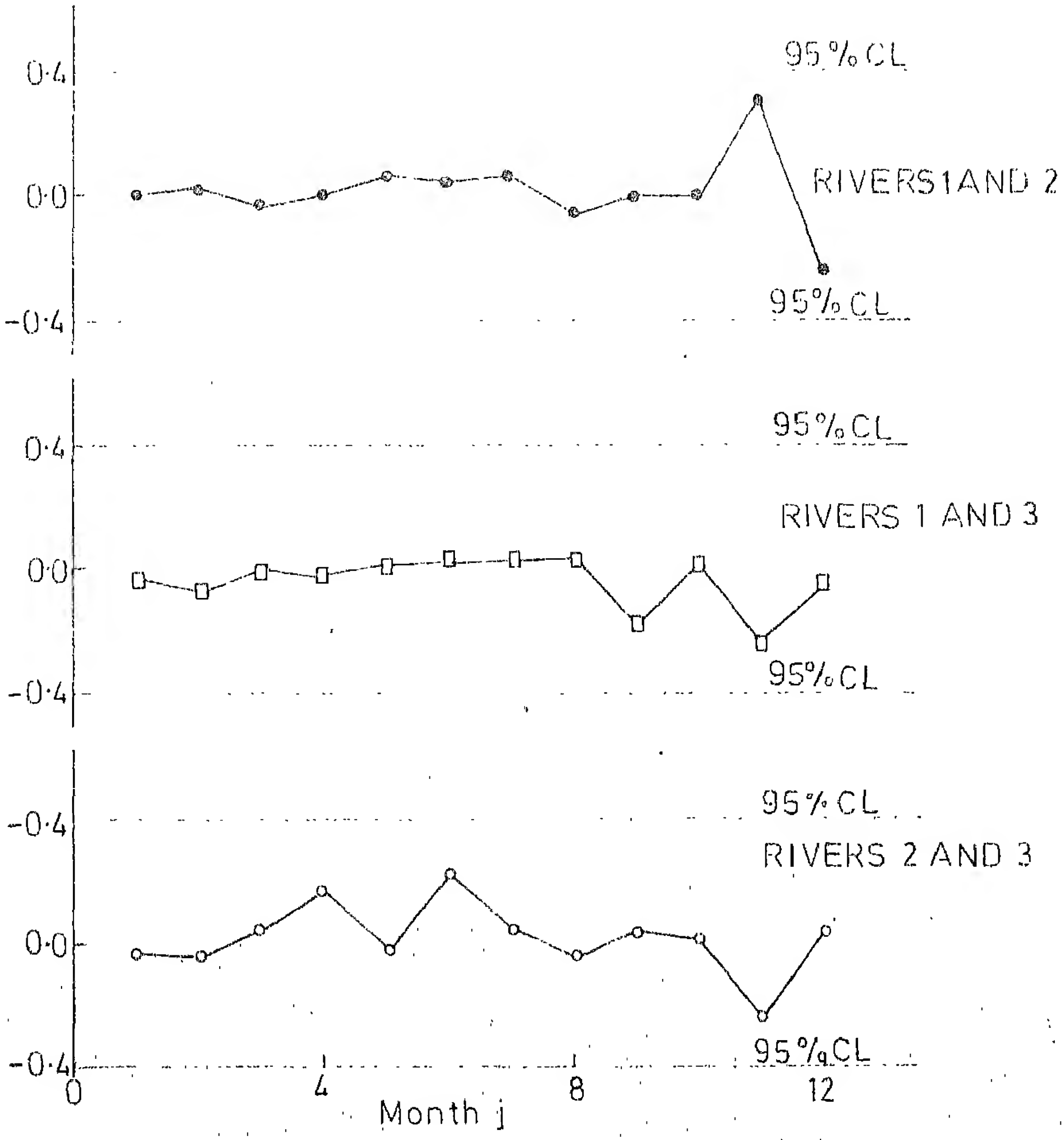
Independence among the time series: The crosscorrelation coefficients between any pair of the multivariate residual time series are estimated for each season. The results for the monthly series (MLS) are given in Fig. 5.12. They may be compared with the corresponding results for the stationary model (Fig. 5.6). For the monthly series, the crosscorrelation is statistically not significant at 95% confidence level for the nonstationary model while it is significant for about 10%



G·5·10 SEASONAL SERIAL CORRELATION COEFFICIENTS OF MULTIVARIATE MONTHLY RESIDUALS (MLS) NONSTATIONARY MODEL



5·11. SEASONAL SERIAL CORRELATION COEFFICIENTS OF MULTIVARIATE TENDAILY RESIDUALS (MLS) NONSTATIONARY MODEL



G. 5·12 SEASONAL CROSS CORRELATION COEFFICIENTS OF MULTIVARIATE MONTHLY RESIDUALS (MLS) NONSTATIONARY MODEL

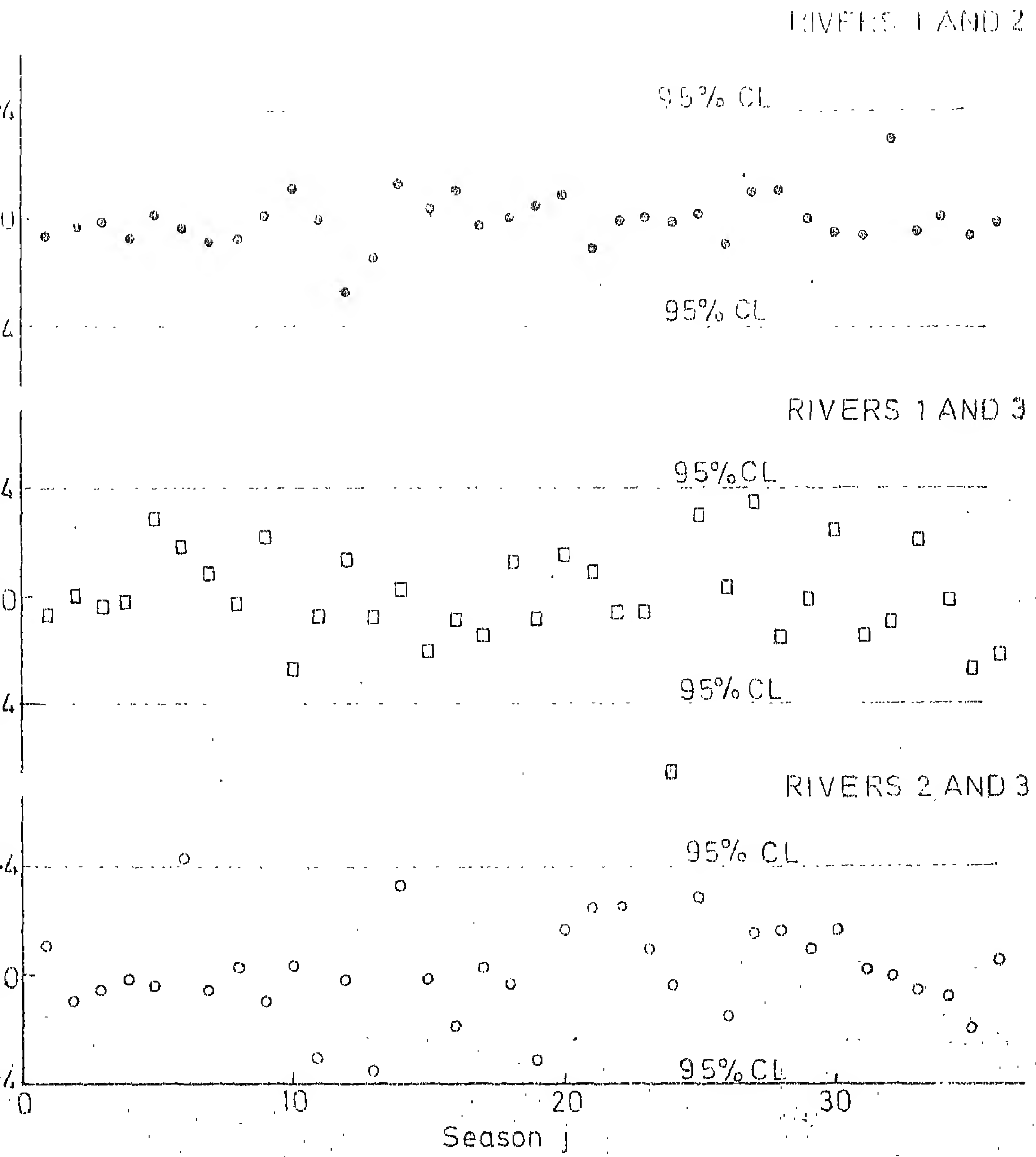
of the cases of the stationary model. For the tendaily series (Fig. 5.13) there is a significant reduction (to almost one) in the number of seasons with significant seasonwise crosscorrelation. The nonstationary model is found to be superior to the stationary model for the monthly and tendaily series and hence is to be preferred.

5.4 Conclusions:

On the basis of the multivariate analysis the following conclusions are derived:

i) A stationary multivariate first order AR model can be fitted to the annual data. The results show that the residuals are not correlated within themselves, among themselves and seasonally. Further they have a nearly normal distribution.

ii) A stationary multivariate zeroeth order AR model can be fitted to the monthly and tendaily data. The fittin of the stationary model involves the estimation of a large number of parameters, but they are estimated generally from a large sample of data. The estimates are hence reliable. By decoupling the variances, the parameters are estimated in two phases. Hence the estimates have a reliability comparable to those of the respective univariate and multivariate models. The residuals of the stationary model are



5·13 SEASONAL CROSSCORRELATION COEFFICIENTS
MULTIVARIATE TENDAILY RESIDUALS (MLS)
NONSTATIONARY MODEL

not serially correlated and do not exhibit significant coherence. However a significant seasonwise crosscorrelation between the residuals is indicated in the case of 10 to 15% of the seasons. Since this violates the assumption of independence of residuals, a stationary model is not considered satisfactory.

iii) A nonstationary multivariate first order AR model can be fitted to the monthly data. For nonstationary modelling, the number of parameters estimated is quite large. For example, for the case studied, for the monthly data, the sample consists of 900 observations and the total number of parameters estimated are 252 including mean, standard deviation and autoregressive coefficient for each river during each season and three elements of $[M_0]$ and 9 elements of $[M_{-1}]$ for each season. Since the parameters are similar to those of the annual series and the sample sizes are also comparable, the accuracy of the results may be considered to be comparable to those of the annual series. The results show that the residuals are not correlated within themselves, among themselves and seasonally. Further they have a nearly normal distribution. So the nonstationary multivariate model seems preferable to the stationary model.

iv) In the case of tendaily data, it is not possible to fit the nonstationary multivariate first order AR model for

4 out of 36 seasons. This is because $[D^j][D^j]^T$ for these seasons is not positive definite. In such cases, it is necessary to either use other suitable models or to use special procedures to take care of the inconsistencies. While a zeroeth order multivariate model may be fitted in such cases, the $[M^j_{-1}]$ matrix is not preserved and this may lead to inconsistency of the model taken as a whole.

6. FLOW FORECASTING AND ESTIMATION

6.1 General

A stochastic model can be used to infer the range of future values based on past records. Forecasting consists in deriving the conditional distribution of the future observations from the past observations. It includes the determination of the characteristics of that distribution, viz., the conditional expectation (mean) referred to herein as the 'forecast' of any future observation and its variability about the forecast. It can be shown that the conditional expectation has the property of being the minimum mean square error forecast. When estimates of the present or past value of one variable is to be made, on the basis of earlier records of itself only or with concurrent and past records of other variables as well as in fitting up of voids in data series, it is referred to as estimation.

Consider a process $x(t)$. A n-step ahead (lead time n) forecast implies the determination of $x(t+n)$ given the present and past values of the process, $x(t)$, $x(t-1)$, $x(t-2)$ etc. in the case of univariate forecasting and corresponding values of related time series in the case of multivariate forecasting. Hence the forecast $x_f(t+n)$ is given by

$E[x(t+n) \cdot x(t), x(t-1), \ldots]$. It is assumed that the present and past values of the time series are available as observations. If $n = 1$, the forecast is referred to as the one-step-ahead forecast. The error of a forecast is given by the difference between the forecasts and actual observations. It is generally assumed that the error is normally distributed. Hence confidence limits for the forecasts can also be specified. The confidence limits express the uncertainity in forecasts and the wider apart they are, the less reliable is the forecast. In addition, since the distant future is more uncertain than the immediate, it will be seen that the confidence interval increases with the lead time.

In the case of a pure random process, the future values do not depend on the present or past values and hence there is no advantage in knowing them. On the other hand, in the case of serially and mutually dependent processes, the deterministic component due to internal and external dependence can be estimated from the present and past values of the time series and the error or estimate is correspondingly reduced. Hence by using the univariate and multivariate time series models, it is possible to forecast the future values with smaller forecasting errors for a given level of confidence. In this study, only one-step-ahead forecasting

is considered. Furthermore, forecasting is done only for the annual and monthly series.

Forecasting generally involves (i) the identification and/or postulation of a model (ii) estimation of the model parameters (iii) validation of the assumed model and, if necessary, repostulation of a different model for validation (iv) adoption of the model for forecasting and (v) comparison of forecasts with observations. Steps (i) to (iii) have already been done in the earlier chapters and the results are used for steps (iv) and (v) in this chapter.

6.2 Forecasting Models

6.2.1 Univariate models

The following univariate time series models have been adopted (Chapter 4) for the normalised standardised series.

i) Annual series

$$x(t) = \varepsilon(t) \tag{6.1}$$

ii) Monthly series:

$$x(t) = \phi_1^j \, x(t-1) + \varepsilon(t) \tag{6.2}$$

when standardisation is done by MLE; and

$$x(t) = \phi_0^j + \phi_1^j \, x(t-1) + \varepsilon(t) \tag{6.3}$$

when standardisation is done by MM and MLS, with $\bar{\varepsilon}^j = -\phi_0^j$.

$$x_f(t) = E[x(t) \mid x(t-1), x(t-2), \ldots]$$

is as follows:

Annual series:

$$x_f(t) = 0 \tag{6.4}$$

Monthly series: For standardisation by MLE

$$x_f(t) = \phi_1^j \, x(t-1) \tag{6.5}$$

and for standardisation by MM and MLS

$$x_f(t) = \phi_0^j + \phi_1^j \, x(t-1) \tag{6.6}$$

The standard error of the forecast is given by s_ε^j, the standard deviation of $\varepsilon(t)$ for season j. The $x_f(t)$ values may then be destandardised to yield $y_f(t)$ values.

Annual series:

$$y_f(t) = x_f(t) \, s_y + \bar{y}, \text{ and,} \tag{6.7}$$

$$Q_f(t) = y_f(t) \tag{6.8}$$

Assuming normality of errors the 95% confidence limits are $y_f(t) \pm 1.96 \; s_y$.

Monthly series: Destandardisation of $x_f(t)$ into $y_f(t)$ gives

$$y_f(t) = x_f(t) \, s_y^j + \bar{y}^j \tag{6.9}$$

The corresponding standard error and 95% confidence limits are, respectively,

$$SE_y^j = s^j \; s_y^j \text{ and} \tag{6.10}$$

$$CL = y_f(t) \pm 1.96 \; SE_y^j \tag{6.11}$$

The forecasts $y_f(t)$ can be transformed back to the original domain using Eq. 3.5 and 3.6.

6.2.2 Decoupled multivariate models

The model adopted consists of univariate models for the normalised and standardised data at all the sites to prewhiten the time series and a nonstationary multivariate first order AR model relating the standardised residuals of the univariate models, viz.,

$$\{X(t)\} = \sum_{i=1}^{p} [\phi_i^j] \{X(t-i)\} + \{\varepsilon(t)\} \tag{5.2}$$

$$\{E(t)\} = \{\varepsilon(t)/s_\varepsilon^j\} \text{ and} \tag{5.3}$$

$$\{E(t)\} = [C^j] \{E(t-1)\} + [D^j] \{\eta(t)\} \tag{5.5}$$

If MLS is used for standardisation, the above expressions are to be modified as indicated in Subsec. 6.2.1.

Combining the above equations,

$$\{X(t)\} = \sum_{i=1}^{p} [\phi_i^j] \{X(t-i)\} + [s_\varepsilon^j][C^j] \{E(t-1)\} + [s_\varepsilon^j][D^j] \{\eta(t)\} \tag{6.13}$$

Further, as $\{Y(t)\} = [s_y^j]\{X(t)\} + \{\bar{Y}^j\}$

$$\{Y(t)\} = \{\bar{Y}^j\} + [s_y^j] \sum_{i=1}^{p} [\phi_i^j] \{X(t-i)\}$$

$$+ [s_y^j][s_\varepsilon^j][C^j] \{E(t-1)\}$$

$$+ [s_y^j][s_\varepsilon^j][D^j] \{\eta(t)\} \qquad (6.14)$$

In the above, $[\phi_i^j]$, $[s_y^j]$ and $[s_\varepsilon^j]$ are diagonal matrices respectively representing the univariate i-th order AR coefficient, standard deviation of the transformed flow and the residual variance for the j-th reason. Without loss of generality, it is assumed that the variable to be forecast or estimated is the last or K-th variable of the K X 1 vector $\{X(t)\}$. In multivariate modelling, all flow data preceding the current flow at all the sites are known. Hence the first 3 terms on the right hand side of Eq. 6.14 represent the deterministic multivariate components for forecast and the last term, the random component for forecast. The expected value of one-step-ahead forecast for the K-th site at time t is given by,

$$x_f(t) = \sum_{i=1}^{p} \phi_i^j (K,K)\, x_K(t-i) + s_\varepsilon^j(K,K) \{C_K^j\}\{E(t-1)\} \qquad (6.15)$$

where $\{C_K^j\}$ refers to the K-th row of the $[C^j]$ matrix, viz.,

$$\{C_K^j\} = \{C^j(K,1),\ C^j(K,2) \ldots C^j(K,K)\}$$

and $x_K(t)$ is x(t) for the K-th site.

The standard error of forecast is given by,

$$SE^j_{x_f} = s^j_\varepsilon\ (K,K)\ \left\{\sum_{n=1}^{K} [d^j(K,n)^2]\right\}^{\frac{1}{2}} \qquad (6.16)$$

where $d^j(K, n)$ is the (K, n)-th element of the $[D^j]$ matrix.

Annual series: A stationary multivariate first order AR model has been fitted. As the univariate model is a pure random series, the ϕ terms are zero. The forecast, its standard error and 95% confidence limits are given respectively by

$$x_f(t) = s_\varepsilon(K,K)\ \{C_K\}\ \{E(t-1)\} \qquad (6.17)$$

and

$$SE_{x_f} = s_\varepsilon\ (K,K)\ \left\{\sum_{n=1}^{K} d^2\ (K,n)\right\}^{\frac{1}{2}} \text{ and}$$

$$CL = x_f(t)\ \pm 1.96\ SE_{x_f} \qquad (6.18)$$

Monthly series: For the stationary multivariate model for the monthly series, the [C] matrix is not significantly different from zero. Forecasting by the multivariate model, then, is the same as that by the univariate model. (See Subsec. 6.2.1). For the nonstationary multivariate model, the forecast and its standard error are given by

$$x_f(t) = \phi^j_1\ (K,K)\ x_K(t-1) + s^j_\varepsilon(K,K)\ \{C^j_K\}\{E(t-1)\}$$

and

$$SE^j_{x_f} = s^j_\varepsilon(K,K)\ [\sum_{n=1}^{K} (d^j(K,\ n))^2]^{\frac{1}{2}} \qquad (6.19)$$

6.2.3 Forecasting

Using the models developed in, and the results of, the previous chapters, one-step-ahead forecasts were made for station 3 for the annual and monthly series. Univariate forecasting of annual flows requires no previous data. Multivariate forecasts of annual flows and multivariate forecasts of monthly flows using the nonstationary model require the present value of the variable at all the sites. Using the methods described in Subsec. 6.2.2, the forecast, its standard error and the 95% confidence limits were calculated.

Annual series: The forecasts were made for 5 years using the univariate model and the multivariate model. They are shown in Fig. 6.1 along with the 95% confidence limits on the forecasts. The standard error of forecast for the multivariate model is 0.86 as against unity for the univariate model. The multivariate forecasts are centred around points other than the mean which corresponds to univariate forecasts, and the confidence limits for the multivariate model are narrower than those for the univariate model. The actual observations lie within the 95% confidence limits. It may be inferred that multivariate forecasts are better than univariate forecasts.

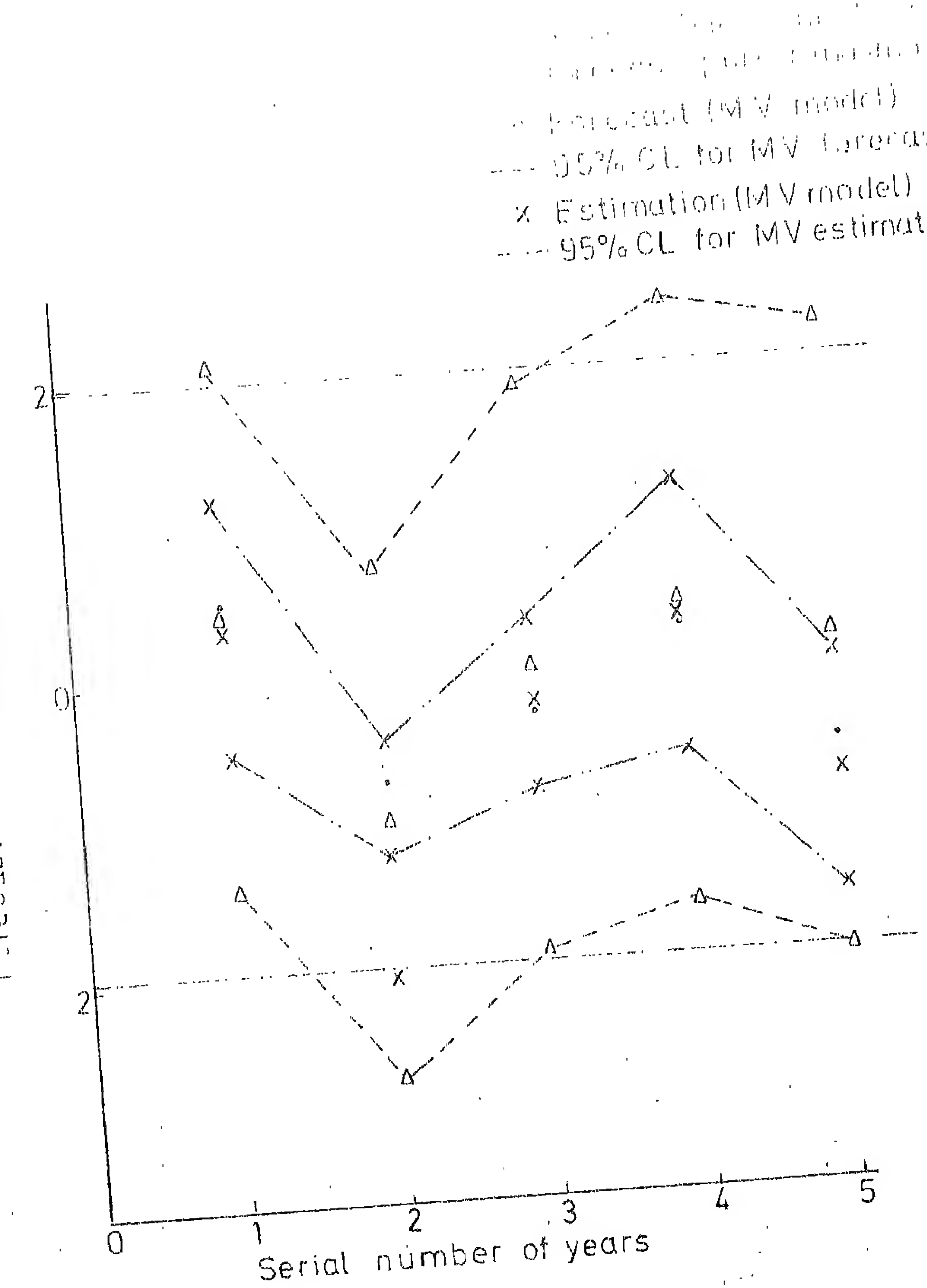


FIG. 6.1 FORECASTS, ESTIMATES AND THEIR CONFIDENCE LIMITS

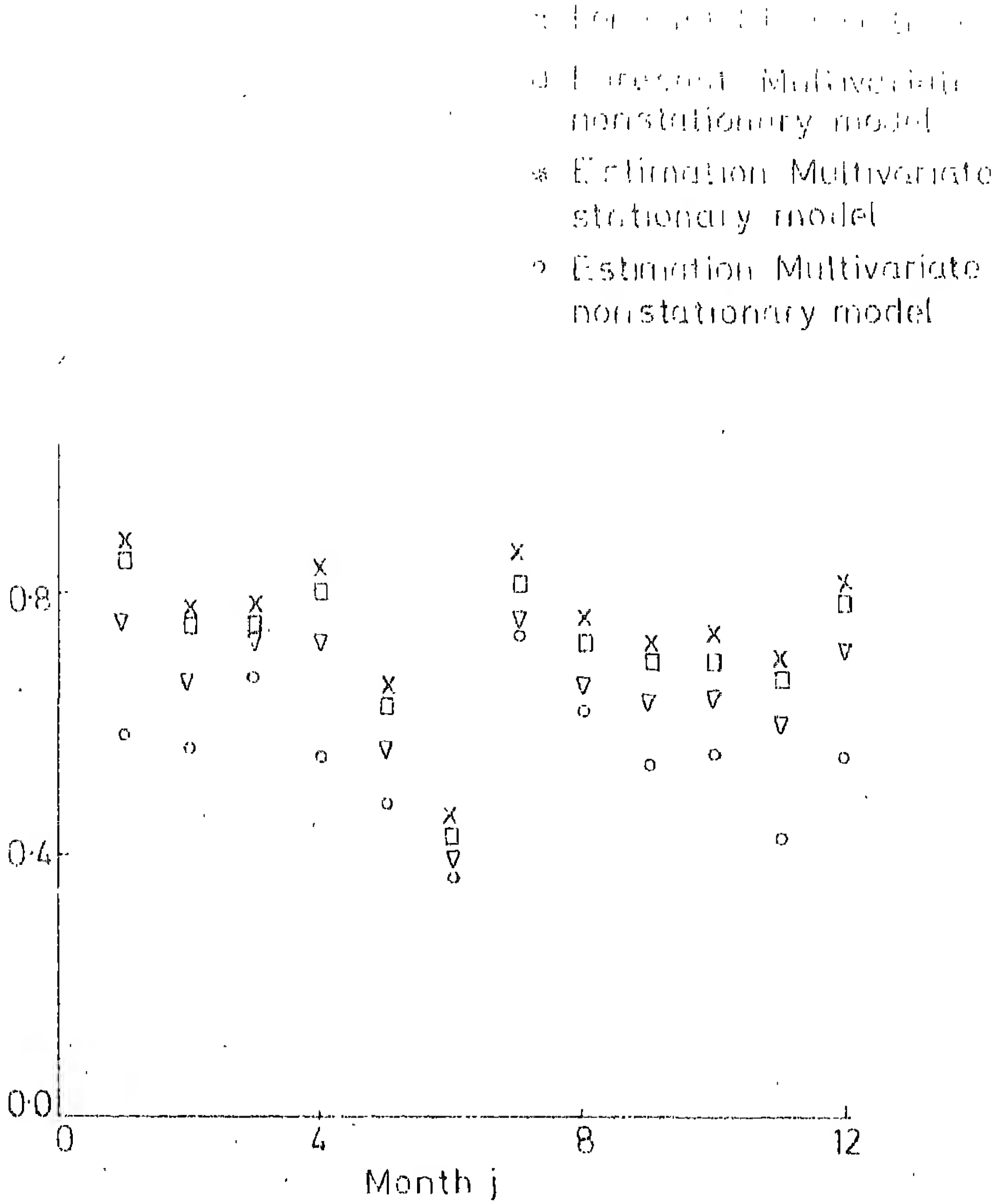
Monthly series: The stationary multivariate model fitted to theunivariate residuals is a zeroeth order AR model and so the forecasts of the stationary model are the same as that for the univariate model. Using the historical record for one year, forecasts for months 2 to 12 were made respectively from the observations for months 1 to 11. The standard errors of forecasts are shown in Fig. 6.2 and the forecasts with their respective 95 % confidence limits are shown in Fig. 6.3.

The univariate forecasts and multivariate forecasts differ from the mean. The standard error for the univariate forecast is less than the standard deviation of the variable and, in turn, the standard error of the multivariate forecast is less than that of the univariate forecast. It is seen that the forecast for the fifth month lies outside the confidence band for any method of forecast and hence seems to be a rare event. The forecasts for other methods are within the 95 % confidence limits.

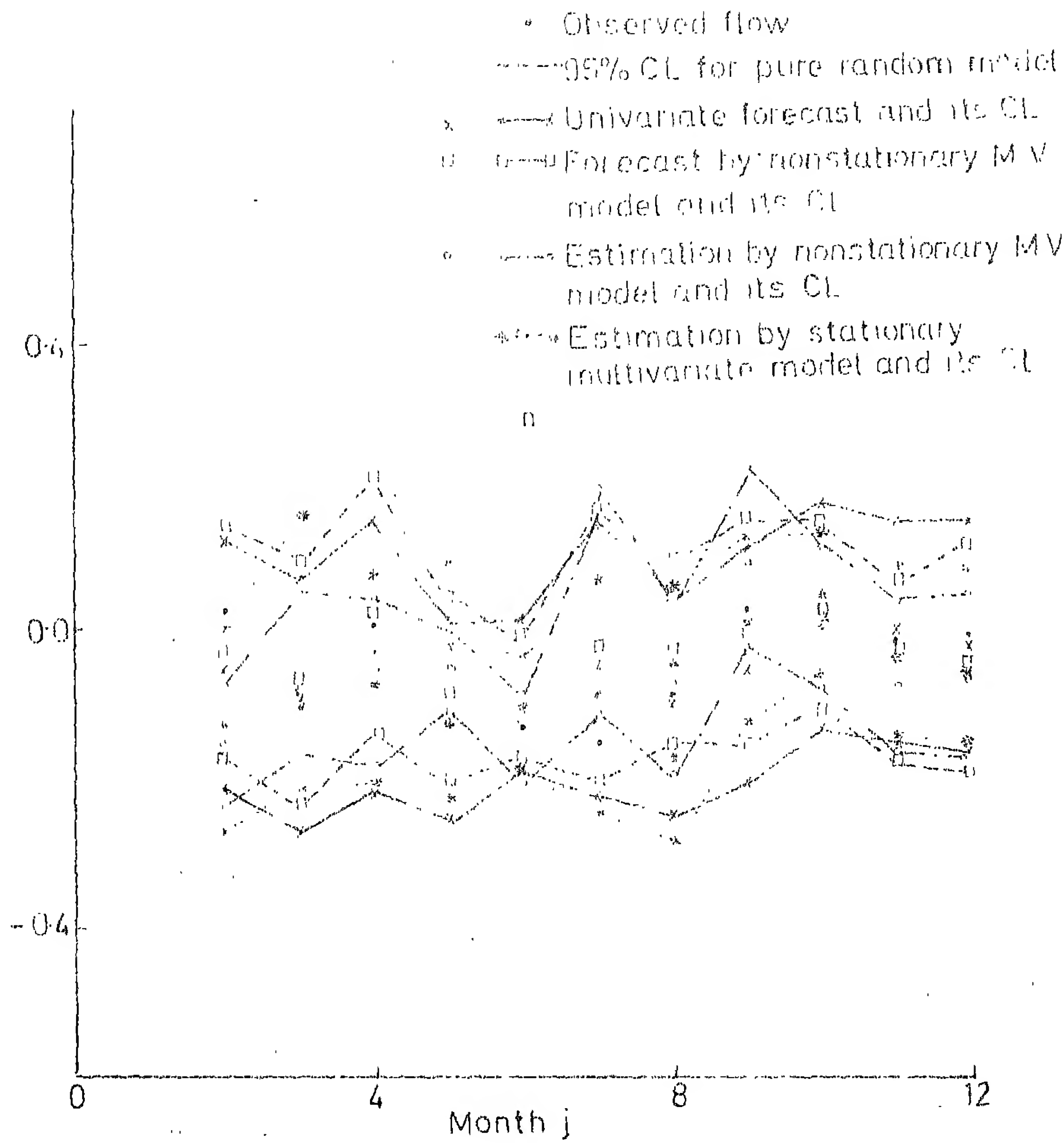
6.3 Estimation

6.3.1 Models for estimation

Univariate estimation and univariate forecasting are identical. In multivariate estimation, in addition to the past values of itself and other stations, current values of



FIG·6·2 STANDARD ERRORS OF FORECASTS AND ESTIMATES (MONTHLY SERIES)



?·6·3 FORECASTS, ESTIMATES AND THEIR CONFIDENCE LIMITS (MONTHLY SERIES)

other stations are also used. The estimate herein made for station K, is given by,

$$x_{est}(t) = s_\varepsilon(K,K) \left[\{C_K\}\{E(t-1)\} + \sum_{n=1}^{K-1} d(K,n)\, \eta_n(t)\right] \tag{6.20}$$

for the annual series, and

$$x_{est}(t) = \phi_1^j(K,K)\, x_K(t-1) + s_\varepsilon^j(K,K)\left[\{C_K^j\}\{E(t-1)\} + \sum_{n=1}^{K-1} d^j(K,n)\, \eta_n(t)\right] \tag{6.21}$$

for the monthly series, where $\eta_n(t)$ is the n-th element of the $\eta(t)$ vector. The corresponding standard errors of estimate are given respectively by

$$SE_{x_{est}} = s_\varepsilon(K,K)\, d(K,K) \tag{6.22}$$

for the annual series, and

$$SE^j_{x_{est}} = s^j_\varepsilon(K,K)\, d^j(K,K) \tag{6.23}$$

for the monthly series.

6.3.2 Results of estimation

The normalised standardised flow in station 3 is estimated using the equations given in Subsec. 6.3.1. Past and current data at stations 1 and 2 as well as the past data at station 3 are used in the estimation.

Annual series: The univariate estimate which is identical to the forecast is the mean flow whereas the multivariate estimate differs from the univariate estimate and the multivariate forecast. Furthermore, the standard error of multivariate estimation is 0.42 which is smaller than that of the multivariate forecast viz., 0.86 which in turn is smaller than that of the univariate forecast of unity. The forecasts and estimates as well as their 95% confidence levels are shown in Fig. 6.1 for a period of 5 years. The actual observations lie within the confidence intervals. Hence for annual series, multivariate estimation is to be preferred to forecasting.

Monthly series: The univariate estimate is the same as the univariate forecast. The standard errors of univariate stationary multivariate and nonstationary multivariate estimates are shown in Fig. 6.2 along with the standard errors of forecasts. They indicate a progressive decrease in the standard error in the following order: univariate forecast or estimation, nonstationary multivariate forecast, stationary multivariate estimation and nonstationary multivariate estimation. The data for the same period as in the case of forecasting were used for estimation. The multivariate estimates and their 95% confidence limits for stationary and nonstationary models are shown in Fig. 6.3. It may be noted that the

estimates are different and they, in turn, differ from the forecasts. It isseen that for month 5, the actual flow lies outside all confidence bands and is a rare event.

For month 2, the actual flow is outside the confidence limits for stationary and nonstationary multivariate estimation and for month 7, it is outside the confidence limits for nonstationary multivariate estimation only. Whether these are due to sampling or modelling error needs further investigation. However because of smaller standard error of estimate associated with the nonstationary multivariate models, they may be preferred to other models.

6.4 Conclusions

i) The decoupled model can be used for one-step-ahead streamflow forecasting as well as for filling-in of gaps in data using data of streamflows at adjacent stations.

ii) The annual data indicate that a multivariate model is useful for forecasting in comparison to a univariate model and that multivariate estimation is preferable to multivariate forecasting.

iii) Forecasts and estimates have been made for 11 months only. The forecasts indicate the general superiority of the nonstationary model over others.

iv) The multivariate estimates have a narrower confidence band than others and for the limited sample used for testing, sometimes result in the actual occurrence being outside the confidence band. Whether this is due to sampling errors needs further investigation.

7. SUMMARY, CONCLUSIONS AND SUGGESTIONS FOR FURTHER STUDY

7.1 Summary

The design and operation of a water resources system should be based on a clear understanding of the processes affecting them which in turn depend on proper analysis and interpretation of data. Particularly important are the streamflow data which exhibit complex stochastic characteristics. They include: 1. a nonstationary behaviour in terms of the seasonal variation of the mean, standard deviation, serial dependence, etc.; 2. trends due to natural or human influences; 3. periodic or cyclic components; 4. autoregression or persistence within the time series; 5. moving average or external correlation between random components at the site; and 6. correlation in space or among time series. Generally factors 1 to 5 are studied in the univariate modelling of a time series, say, the streamflow at a site and factor 6 is studied in the multivariate modelling of several time series. Knowing the dependence or independence of streamflows at different sites, it is possible to design and operate multireservoir systems by taking advantage of the interdependence.

Mathematical modelling of streamflow at a site is itself fairly complicated. Modelling of multisite streamflows is further complicated by the fact that the multivariate streamflow variability may be attributed to i) the within-the-station variations in terms of factors 1 to 5, and ii) the among-the-stations variations in terms of the crosscorrelations between the streamflows at different sites. Univariate models in hydrology deal with the former and multivariate models generally consider both simultaneously. It seems possible to decouple the variability due to these two classes so that the within-the-station variability can be first estimated by univariate modelling and used for identifying the serially independent random components at each site; and then these random components can be related through a multivariate model to account for the inter-station correlations and the among the stations variability. Such an approach seems to be advantageous because of the following:

1) The expertise developed in the past in univariate time series modelling can be beneficially used.

2) As the serially independent and normally distributed variates are used in the multivariate model, a better performance of the model can be expected.

3) Since the parameters are estimated in stages, the number of parameters estimated in each stage is comparable

to that of univariate and multivariate model respectively and the problem of simultaneously estimating all the parameters is avoided.

4) Where data lengths vary from station to station, all the available data can be used in the estimation of parameters and decoupling may eliminate errors and bias that may be present in the simultaneous estimation of parameters.

5) Decoupling of spatial and temporal variations has been suggested recently by Yevjevich, Mejia and Iturbe, and Yevjevich and Karplus. Rather than using simple spectral and correlation procedures to represent spatial variability, the use of multivariate models is suggested in this study. This is similar to the approaches of Frost and Clarke and Yevjevich.

6) Decoupled multivariate models seem to be very powerful tools in representing complex multiple interrelated time series.

Without loss of generality, the study was restricted to three sites in a river basin in North India for which data were available and to a 25 year concurrent record of streamflow. Data series considered in this study include i) the tendaily historical data series; ii) the monthly data series calculated from the tendaily series as the average of the

three tendaily values in each month; and iii) the annual series, being the average of the monthly values in each year. Annual data were normally distributed; but monthly and ten daily series were normalised by logarithmic transformation. The data series were standardised to a mean zero, unit variance series by means of the sample means and standard deviations for each of the periods. Using ARIMA models and a nonlinear least squares regression approach, univariate models were fitted to the standardised normalised time series at each site. The results indicated autoregressive models of orders zero, one and three respectively for the annual, monthly and tendaily series. The residuals were calculated and tested for serial independence. Time domain and frequency domain analyses indicated serial independence when the full series was considered; but, for monthly and ten daily series the correlation coefficients between the residuals of adjacent periods were found to be significant in several cases. This indicated a seasonal variation of persistence and so nonstationary autoregressive models of the same order as before but with seasonally varying parameters were fitted to the monthly and tendaily series. The residuals from the nonstationary models were found to be serially uncorrelated seasonally and as a whole. The univariate residuals were also generally normally distributed. The normally distributed serially independent residual

series are generally referred to as 'prewhitened' series and the process of identifying and separating such residuals is referred to as 'prewhitening'. Hence the first stage in decoupled multivariate modelling involves 'prewhitening' the univariate data series.

The second stage consists in fitting a multivariate time series model to the univariate residuals to take into account the among-the-series variation. Using currently available procedures (Matalas, Young and Pisano), a stationary first order autoregressive multivariate model was fitted to the univariate residual series. The multivariate residuals were evaluated and tested for independence serially and with one another. When considered as a whole, the series were found to be independent serially and of one another; but, monthly and tendaily series showed significant crosscorrelation on a seasonal basis in several periods. This suggested a nonstationary multivariate model with seasonally varying parameters. For the tendaily series, for four seasons it was not possible to fit nonstationary first order autoregressive multivariate models using the procedures adopted for stationary modelling. This is because of inconsistencies in the sample correlation matrices and the problem was similar to cases mentioned by Matalas and Wallis. Zeroeth order models were fitted to the tendaily series in

series are generally referred to as 'prewhitened' series and the process of identifying and separating such residuals is referred to as 'prewhitening'. Hence the first stage in decoupled multivariate modelling involves 'prewhitening' the univariate data series.

The second stage consists in fitting a multivariate time series model to the univariate residuals to take into account the among-the-series variation. Using currently available procedures (Matalas, Young and Pisano), a stationary first order autoregressive multivariate model was fitted to the univariate residual series. The multivariate residuals were evaluated and tested for independence serially and with one another. When considered as a whole, the series were found to be independent serially and of one another; but, monthly and tendaily series showed significant crosscorrelation on a seasonal basis in several periods. This suggested a nonstationary multivariate model with seasonally varying parameters. For the tendaily series, for four seasons it was not possible to fit nonstationary first order autoregressive multivariate models using the procedures adopted for stationary modelling. This is because of inconsistencies in the sample correlation matrices and the problem was similar to cases mentioned by Matalas and Wallis. Zeroeth order models were fitted to the tendaily series in

such cases. Development of procedures for fitting first order models and for estimation of parameters in such cases is beyond the scope of this study. Monthly data did not exhibit any such problems. The multivariate residuals from the nonstationary model were estimated and when tested statistically, they were found to be independent serially and of one another, seasonally and as a whole.

It is possible to recouple the multivariate and univariate models and also use inverse transformations to denormalise and destandardise them in the proper order so that the complex relationships between the actual flows can bederived in terms of the parameters estimated earlier. These relationships can be used to predict missing data, if any; forecast future values of the streamflows at any site; and generate multisite streamflow data for simulation of complex water resources systems.

One step ahead forecasting and predictions were considered in this study using the multivariate model developed earlier. It was seen that generally multivariate estimation and forecasting leads to a better definition of the expected value and confidence levels of the variable concerned than in the case of univariate models.

7.2 Conclusions

On the basis of this study, the following conclusions can be made:

i) Decoupling facilitates the use of simpler component models in the modelling of complex stochastic processes, and leads £to a better understanding of the temporal and spatial dependence between streamflows at different sites.

ii) For the standardised annual series, a zeroeth order univariate autoregressive model and a first order multivariate autoregressive model are indicated. For the normalised, standardised monthly series, a nonstationary first order autoregressive univariate and multivariate model are indicated. Tendaily data after normalisation and standardisation, can be represented generally by a third order autoregressive nonstationary univariate model and a first order autoregressive nonstationary multivariate model. For some seasons, the sample correlation coefficients for the non-stationary model were inconsistent with the assumption of a multivariate first order autoregressive model. In such cases, a zeroeth order multivariate model was fitted to the tendaily series.

iii) Multivariate models can be used for the estimation of missing values, and for one step-ahead forecasting. They generally lead to a better definition of the conditional

expectation and a smaller standard deviation of the forecast or estimate.

iv) Data can be generated on the basis of the decoupled multivariate model and used in simulation analysis of multi-reservoir systems. Decoupling is a very powerful tool in the mathematical modelling of interrelated time series and the procedure developed herein can be used for modelling other multivariate processes.

7.3 Suggestions for Further Study

Based on the results of the study, the following suggestions are made for future work in this area:

i) Decoupled multivariate models have been used for representing multisite streamflows in three rivers of North India. The decoupled model consists of nonstationary univariate AR models and zeroeth or first order multivariate AR models relating the univariate residuals. It seems possible to develop other models in a decoupled framework. Investigations should include resolution of inconsistencies between data and assumed models. Furthermore, the procedure may be applied to other regions of India and abroad in order that regional space-time correlations can be identified and defined.

ii)

ii) Decoupled multivariate models have already been used to represent input-output relationships in other areas of engineering. The component processes of hydrologic cycle have been represented through input-output models. Hence it seems possible to represent input-output models in hydrology through decoupled multivariate models.

iii) Generally, in the representation of complex processes, simple models with parsimony in parameters are preferred. A decoupled model with simple components may hence be preferred to more complicated nondecoupled models. The results of the study indicate that the interstation correlation is nonstationary in at least some seasons and hence a nonstationary multivariate model is required on conceptual grounds. It seems necessary to use simulation models and generated data based on stationary and nonstationary models to justify on pragmatic grounds the need for more complicated nonstationary multivariate modelling.

REFERENCES

1. Anderson, T.W., An Introduction to Multivariate Statistical Analysis, John Wiley and Sons, New York, 1958.

2. Beard, L.R., 'Use of Interrelated Records to Simulate Streamflows', Journal of the Hydraulics Division, ASCE, Vol. 91, No. HY5, 1965, pp. 13-22.

3. Box, G.E.P., and Jenkins, G.M., Time Series Analysis, Forecasting and Control, Holden-Day Publishing Co., San Franscisco, 1970.

4. Carlson, R.F., McCormick, A.J.A., and Watts, D.G., 'Application of Linear Random Models to Four Annual Streamflow Series', Water Resources Research, Vol.9, No.4, 1970, pp. 1070-1078.

5. Chow, V.T., Handbook of Applied Hydrology, McGraw Hill Book Co., New York, 1964.

6. Clarke, R.T., Mathematical Models in Hydrology, Food and Agricultural Organisation, Rome, 1973.

7. Cox, D.R., and Miller, H.D., The Theory of Stochastic Processes, John Wiley and Sons, New York, 1965.

8. Crosby, D.S., and Maddock III, T., 'Estimating Coefficients of a Flow Generator for Monotone Samples of Data,' Water Resources Research, Vol. 6, No. 4, 1970, pp. 1079-1086.

9. Feller, W., 'The Asymptotic Distribution of the Range of Sums of Independent Random Variables', Annals of Mathematical Statistics, Vol. 22, 1951, pp. 427-430.

10. Fiering, M.B., 'A Multivariate Technique for Synthetic Hydrology', Journal of the Hydraulics Division, ASCE, Vol. 90, No. HY5, 1964, pp. 43-60.

11. Fiering, M.B., Streamflow Synthesis, Harvard University Press, Cambridge, Massachusetts, 1967.

12. Fiering, M.B., 'Schemes for Handling Inconsistent Matrices', Water Resources Research, Vol. 4, No. 2, 1968, pp. 291-297.

13. Frost, J., and Clarke, R.T., 'Use of Crosscorrelation Between Hydrologic Time Series to Improve Estimates of Lag one Autoregressive Parameters', Water Resources Research, Vol.9, No. 4, 1973, pp. 906-917.

14. Fuller, G.A., 'Inconsistent Matricesin Estimating Ungaged Streamflow Data', Water Resources Research, Vol. 10, No. 6, 1974, pp. 1249-1256.

15. Garcia, L.E., Dawdy, D.R., and Mejia, J.M., 'Long Memory Monthly Streamflow Simulation by A Broken Line Model', Water Resources Research, Vol. 8, No.4, 1972, pp. 1100-1105.

16. Granger, C.W.J., and Hatanka, M., Spectral Analysis of Economic Time Series, Princeton University Press, Princeton, New Jersey, 1964.

17. Hamlin, M., Unpublished Report on the Workshop on Stochastic Hydrology, Symposium on the Application of Mathematical Models in Hydrology, Bratislava, 1975.

18. Hurst, H.E., 'Long Term Storage Capacity of Reservoirs', Transactions, American Society of Civil Engineers, Vol.116, 1951, pp. 770-799.

19. Langbein, W.B., 'Queuing Theory and Reservoir Storage', Proceedings of the Hydraulics Division, ASCE, Vol. 84, No. HY5, 1958, pp. 1811-1831.

20. Linsley, R.K., Kohler, M.A., and Paulhus, J.L.N., Hydrology for Engineers, McGraw Hill Book Co., New York, 1958.

21. Maass, A., Hufschmidt, M.M., Dorfman, R., Thomas, Jr., H.A., Marglin,S.A., and Fair, G.M., The Design of Water Resource Systems, Harvard University Press, Cambridge, Massachusetts, 1962.

22. Mandelbrot, B.B., 'A Fast Fractional Gaussian Noise Generator', Water Resources Research, Vol. 7, No. 3, 1971, pp. 543-553.

23. Mandelbrot, B.B., and Wallis, J.R., 'Noah, Joseph and Operational Hydrology', Water Resources Research, Vol. 4, No. 5, 1968, pp. 909-918.

24. Mandelbrot, B.B., and Wallis, J.R., 'Computer Experiments with Fractional Gaussian Noises', Water Resources Research, Vol. 5, No. 1, 1969, pp. 228-267.

25. Matalas, N.C., 'Mathematical Assessment of Synthetic Hydrology', Water Resources Research, Vol. 3, No. 4, 1967, pp. 931-945.

26. Matalas, N.C., and Wallis, J.R., 'Statistical Properties of Multivariate Fractional Noise Processes', Water Resources Research, Vol. 7, No. 6, 1971, pp. 1460-1468.

27. Matalas, N.C., and Wallis, J.R., 'Correlation Constraints for Generating Processes', Proceedings Symposium on Mathematical Models in Hydrology, Warsaw, 1972.

28. McKerchar, A.I., and Delleur, J.W., 'Stochastic Analysis of Monthly Flow Data - Application to Lower Ohio River Tributaries, Technical Report No. 26, Water Resources Research Center, Purdue University, Lafayette, Indiana, 1972.

29. McMahon, T.A., Codner, G.P. and Philips, C., 'Single and Multisite Operational Hydrology', Nordic Hydrology, Vol. 3, No. 4, 1972, pp. 214-238.

30. Mejia, J.M., Dawdy, D.R., and Nordin, C.F., 'Streamflow Simulation, 3. Broken Line Process and Operational Hydrology', Water Resources Research, Vol. 10, No. 2, 1974(a), pp. 242-245.

31. Mejia, J.M., and Iturbe, I.R., 'On the Synthesis of Random Field Sampling from the Spectrum. An Application to the Generation of Hydrologic Spatial Processes', Water Resources Research, Vol. 10, No. 4, 1974, (b), pp. 705-712.

32. Moran, P.A.P., Theory of Storage, Methuen, London, 1959.

33. Nelson, C.R., Applied Time Series Analysis for Managerial Forecasting, Holden-Day Publishing Co., San Francisco, 1973.

34. Nordenson, T.J., Application of Conceptual Catchment Models to River Forecasting, Technical Note No. 92, World Meteorological Organisation, Geneva, 1969.

45. Wallis, J.R., and Matalas, N.C., 'Small Sample Properties of H and K,' Water Resources Research, Vol. 6, No. 6, 1970, pp. 1583-1594.

46. Wallis, J.R., and Matalas N.C., 'Correlogram Analysis Revisited,' Water Resources Research, Vol. 7, No.6, 1971, pp. 1448-1459.

47. Yevdjevich, V.M., Fluctuations of Wet and Dry Years, Part II, Hydrology Paper No 2, Colorado State University, Fort Collins, Colorado, 1964.

48. Yevjevich, V., Systems Approach to Hydrology, Water Resources Publications, Fort Collins, Colorado, 1971.

49. Yevjevich, V., Probability and Statistics in Hydrology, Water Resources Publications, Fort Collins, Colorado, 1972(a).

50. Yevjevich, V., Stochastic Processes in Hydrology, Water Resources Publications, Fort Collins, Colorado, 1972(b).

51. Yevjevich, V., Generation of Hydrologic Samples. Case Study of the Great Lakes, Hydrology Paper No. 72, Colorado State University, 1975.

52. Yevjevich, V., and Karplus, A.K., Area-Time Structure of the Monthly Precipitation Process, Hydrology Paper No. 64, Colorado State University, Fort Collins, Colorado, 1972.

53. Young, G.K., and Pisano, W.C., 'Operational Hydrology Using Residuals', Journal of the Hydraulics Division, ASCE, Vol. 94, No. HY4, 1968, pp. 909-923.

35. Pentland, R.L., and Cuthbert, D.R., 'Operational Hydrology for Ungaged Streams by the Grid Square Technique', Water Resources Research, Vol. 7, No.2, 1971, pp. 283-291.

36. Quimpo, R.G., 'Stochastic Model of Daily River Flow Sequences', Hydrology Paper No. 18, Colorado State University, Fort Collins, Colorado, 1967.

37. Ramaseshan, S., 'Stochastic Modelling of Storm Precipitation', Nordic Hydrology, Vol. 2, 1971, pp. 109-129.

38. Ramaseshan, S., and Krishnasami, M., 'Decoupled Stochastic Models for Multisite Streamflows', Proceedings, Second World Congress, IWRA, New Delhi, 1975, Vol. 5, pp. 45-50.

39. Rao, A.R., and Delleur, J.W., The Instantaneous Unit Hydrograph, Its Calculation by the Transform Method and Noise Control by Digital Filtering, Technical Report No. 20, Water Resources Research Center, Purdue University, Lafayette, Indiana, 1971.

40. Roesner, L.A., and Yevdjevich, V.M., 'Mathematical Models for the Time Series of Monthly Rainfall and Monthly Runoff', Hydrology Paper No. 16, Colorado State University, Fort Collins, Colorado, 1966.

41. Sariahmed, A., and Kisiel, C.C., 'Synthesis of Sequences of Summer Thunderstorms volumes for the Atterbury Watershed in the Tucson Area', Proceedings, Symposium on Use of Analog and Digital Computers in Hydrology, Tucson, Arizona, 1968, Vol. 2, pp. 439-448.

42. Stidd, C.K., The n-th Root Normal Distribution of Precipitation, Water Resources Research, Vol. 6, No.4, 1970, pp. 1095-1103.

43. Torranin, P., Applicability of Canonical Correlation in Hydrology, Hydrology Paper No. 58, Colorado State University, Fort Collins, Colorado, 1972.

44. Valencia, D., and Schaake, Jr., J.C., 'Disaggregation Processes in Stochastic Hydrology', Water Resources Research, Vol. 9, No. 3, 1973, pp. 580-585.

Date Slip

This book is to be returned on the date last stamped.

CD 6.72.9